Natural England Commissioned Report NECR006

Analysis of vegetation data from different quadrat sizes

First published 26 June 2009

www.naturalengland.org.uk

NATURAL ENGLAND

Introduction

Natural England commission a range of reports from external contractors to provide evidence and advice to assist us in delivering our duties. The views in this report are those of the authors and do not necessarily represent those of Natural England.

Background

There are a huge number of historical datasets containing information on semi-natural vegetation gathered from across the United Kingdom's bioclimatic regions.

These data could provide a valuable opportunity to look at historical vegetation changes in the context of climate change and other environmental drivers. However, the lack of standardisation in the way these data were gathered makes meta-analysis and comparison difficult.

This report results from research commissioned by Natural England in order to investigate the utility of frequency-area curves in standardising vegetation datasets collected using different quadrat sizes. It examines one potential method of standardisation and points the way to possible solutions to the problem.

This research now needs to be reviewed in the context of:

- Analyzing large and varied vegetation data sets.
- Assessing the usefulness of the method against other more standard methods.

If problems inherent in the analytical method can be resolved, it may prove to be a powerful tool for detecting small scale or short-term vegetation changes, for example, in response to long-term climate change.

This report should be cited as:

SCOTT, W.A. & SMART, S.M. 2009. *Analysis of vegetation data from different quadrat sizes*. Natural England Commissioned Report, Number 006.

Acknowledgments

The ESA monitoring programme was funded by MAFF (now Defra) and was developed and implemented by ADAS in consultation with MAFF and the national conservation agencies. The national ESA monitoring programme was co-ordinated by A. J. Hooper, with assistance from J.A. Slater and K.M. Joyce.

The Countryside Survey was funded by Defra and the Natural Environment Research Council (NERC) and was organised and implemented by staff of the Centre for Ecology and Hydrology CEH).

Natural England Project Manager - Clive Bealey, Natural England, 1 Southampton Road, Lyndhurst, Hampshire, SO43 7BU, Tel: 0238 028 6410, Fax: 0238 028 3834, **clive.bealey@naturalengland.org.uk**.

Contractor - W.A. Scott & S.M. Smart, Centre for Ecology and Hydrology, Merlewood Research Station, Windermere Road, Grange-over-Sands, Cumbria, LA11 6JU, **www.ceh.ac.uk**.

Keywords - Data standardisation, long-term monitoring, quadrat size, validation network, vegetation data

Further information
This report can be downloaded from the Natural England website: **www.naturalengland.org.uk**. For information on Natural England publications contact the Natural England Enquiry Service on 0845 600 3078 or e-mail **enquiries@naturalengland.org.uk**.

ISSN 2040-5545

Summary

This report describes the results of a project funded by Natural England to investigate the utility of frequency-area curves in standardising vegetation datasets collected using different quadrat sizes. Techniques for estimating frequency-area curves and for standardising datasets are developed.

Vegetation datasets collected using nested quadrats were made available for analysis by Natural England, the Centre for Ecology and Hydrology (CEH) and Defra/ADAS. These datasets are complimentary, covering between them quadrat areas from 0.004 m^2 to 200 m^2 and a comprehensive range of habitats.

Fourteen different forms of frequency-area curve, were compared. The parameters of the curves have simple interpretations in terms of curve shape and ecological/environmental implications. The range of models studied provided an adequate representation of the majority of species, for all of the datasets, and there is clear evidence that just two of the parameterisations studied are capable of representing the frequency-area relationships of the majority of species. Overall the most suitable model is the logistic *cd* parameterisation but the logistic *acd* parameterisation is necessary for adequate representation of some species, particularly when the number of nests is high and their range extends close to zero size.

The same models were appropriate whether the data represented homogeneous vegetation, such as the individual Natural England sites, or very heterogeneous vegetation such as that from the complete CS dataset.

Accurate identification and representation of a frequency-area curve requires that species be observed in a reasonably large number of quadrats and at a wide range of nest sizes. For identification species to be modelled should be observed in at least twenty plots and over six or more different nest sizes, for prediction presence in ten or more quadrats may be sufficient. The range of nest sizes is also important. Many species exhibit a steep increase in frequency with area before reaching a relatively constant value. If possible the range of nest sizes should cover both parts of the curve**.**

For each of the three dataset, the ordinations obtained from different nest size are very similar suggesting that nest size is not crucial to analysis and that the same information and conclusions will be obtained regardless of which nest size is used. However, datasets compiled from mixed quadrat sizes can give very different results from datasets in which all quadrats are the same size. In the mixed situation quadrat size tends to override the main axes of variation in the data. In addition mixing quadrat sizes can suppress or exaggerate features of the data, with the extent of the distortion depending on the range of quadrat sizes used. Standardisation of datasets for differences in quadrat size is therefore advisable before analysis.

Two methods of standardisation are examined. The first makes a simple adjustment to the frequencies of species in one dataset in order to make the average frequency the same as that of a second dataset. This method has the advantage of being easy to apply, is applicable to all datasets and does not require nested data. It can, in some situations, largely eliminate differences due to varying quadrat size. However, it may also eliminate real differences between datasets. A modification of the method equalises the average frequency of each site.

The second method of standardisation uses frequency-area curves fitted to individual species to predict the frequency of each species at a specified quadrat size, and hence can only be applied to nested or mixed quadrat size data. In all three datasets a substantial proportion of species had insufficient data to fit an accurate frequency-area curve, necessitating an ad hoc method of prediction to be employed for these, and overall the average frequency of the predictions differed markedly from the correct value, particularly at extreme quadrat sizes.

The main conclusion of the project is that frequency-area curves have only a limited utility in standardising vegetation datasets. However recommendations are also made as to the procedure to follow should standardisation via frequency-area curves be attempted.

Contents

Appendices

List of tables

List of figures

1 Introduction

Background to the study: Natural England's Validation Network

1.1 Natural England's Validation Network is a five-year programme designed to quantitatively validate Condition Assessment monitoring on SSSIs in England. Condition Assessment methods involve fairly rapid assessment techniques and assign habitat feature monitoring units to one of five categories of habitat condition (from 'Favourable-maintained' to 'Unfavourable-declining') with two more categories for loss (destroyed and partially destroyed). Quantitative validation will require more traditional, quadrat-based methods for recording botanical composition and physical/environmental attributes.

1.2 For the Validation Network a suite of sites of different 'condition', as assessed by Condition Assessment methods, will be selected for each habitat feature to be studied. Site selection criteria include sites that have had long-term monitoring (of any kind) on them. If the Validation Network is to utilise the historic information from such sites, it will need to be able to compare and, if possible, integrate different types of vegetation data gathered using a wide variety of methods.

1.3 Quantitative vegetation data are usually collected from clearly defined sample plots called quadrats. These are traditionally small (2 x 2 m or less) and square although provision is often made for rectangular quadrats in linear habitats and for larger quadrats in habitats such as woodland. The type and size of quadrat is usually decided subjectively according to the scale of the vegetation and the linearity of the habitat being sampled. Detailed recording methods are also subjectively chosen according to research/monitoring needs. Methods range from presence/absence (reported as frequency for groups of quadrats), through semi-quantitative cover-abundance (for example, Domin) scales to fully quantitative methods such as pin quadrats. In general it is not possible to directly compare the many types of quantitative/semi-quantitative data available. However much of this data can be reduced to presence/absence form and such data can be compared across sites as long as it is corrected for differences in quadrat area.

1.4 The need to correct for quadrat area arises because the recorded frequency of any species varies with quadrat size and the effect of quadrat area on frequency is typically non-linear. A grass species, for example may be recorded with 100% frequency in ten 2 x 2 m quadrats but at only at 50% frequency in 10 1x1m quadrats and 15% frequency in ten 0.5 x 0.5 m quadrats from the same plot. Figure 1 shows such a relationship with the marked points A, B and C equivalent to this example.

1.5 Adjusting datasets to correct for differing quadrat size requires knowledge of species' cumulative frequency-area curves spanning the different sampling scales. Estimating such curves requires data from a range of quadrat sizes such as that provided by nested quadrat data. These estimated curves can then be used to extrapolate to the required "standard" quadrat area for comparison with other datasets. If the modelled/projected frequency-area curves obtained from small or large quadrat areas accurately reflect real curves, then correcting to a standardised quadrat size is not a problem. However, if the errors associated with extrapolation are large the standardised dataset may be of little practical use.

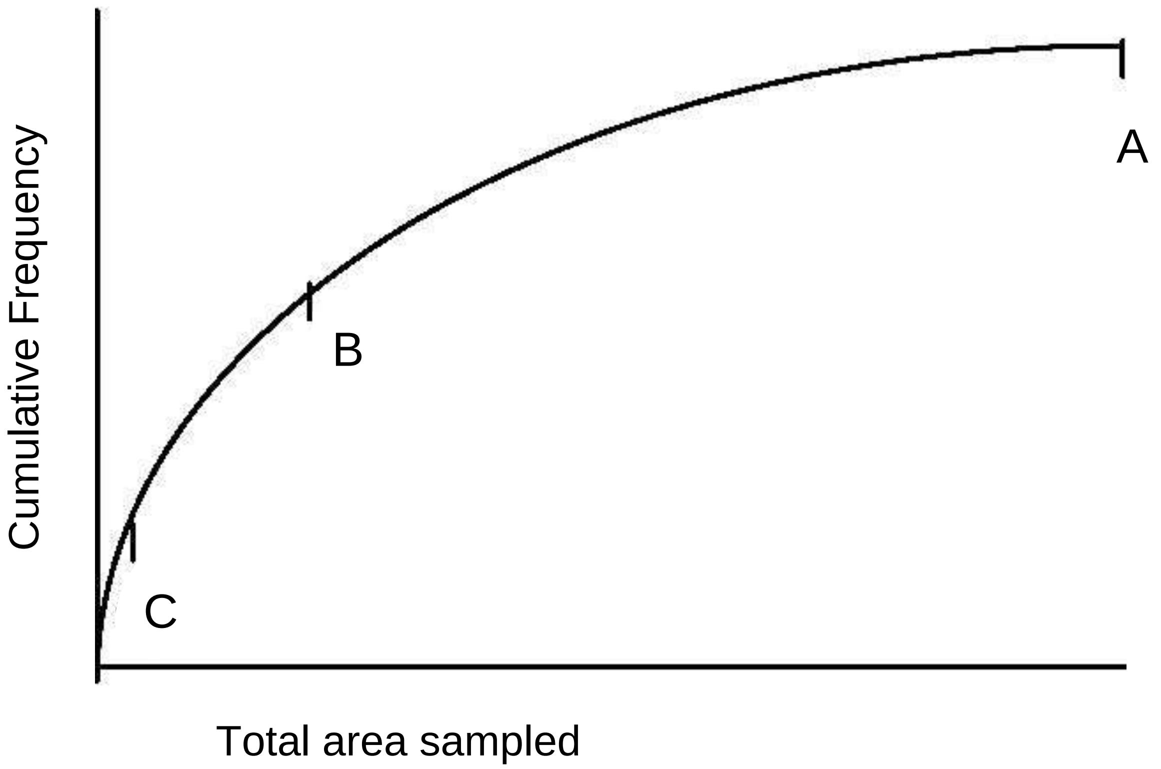


Figure 1 Relationship between sample area and cumulative frequency

Objectives

1.6 The objectives of this project are to develop techniques for estimating frequency-area curves and for standardising datasets collected using different methods, and to make recommendations as to the situations in which comparison of standardised datasets is feasible.

1.7 This report describes the analyses undertaken in order to achieve these objectives, and the application of their results to Natural England's monitoring programmes. The report specifically addresses the following areas:

- Section 4: comparison of alternative models for frequency-area relationships and an assessment of their goodness of fit when used to model actual frequency-area relationships for individual species derived from nested quadrat data.
- Section 5: variation in frequency-area relationships for individual species.
- Section 6: the effect, on analysis at the vegetation community level, of using different quadrat sizes.
- Section 7: standardisation of datasets to a given quadrat size using modelled frequency-area curves, or other methods, and an assessment of the accuracy of the results.

1.8 The datasets used in the project are described in Section 2 and the methods in Section 3, in a form that it is hoped can be understood by staff with medium technical and statistical knowledge. More formal details of the statistical methods involved are given in Appendices 1 and 2.

2 Datasets

2.1 Vegetation datasets collected using nested quadrats were made available for analysis by Natural England, the Centre for Ecology and Hydrology (CEH) and Defra/ADAS. These datasets are complimentary, covering between them quadrat areas from 0.004 m^2 to 200 m^2 and a comprehensive range of habitats. However the recording methods used are not identical in all studies and species are recorded to different levels. For simplicity, the term ‘species’ is used subsequently in this report to refer to both individual species and amalgams. It should also be noted that the analyses carried out during the course of the present study were undertaken with a view to examining and illustrating the effects of differing quadrat size on analysis and possible methods for correcting for this. They are in no way intended to be an opposite or comprehensive analysis of any of the datasets in terms of the rationale for which the datasets were compiled.

Natural England

2.2 Natural England and the other statutory conservation agencies in the UK are monitoring designated sites, including SSSIs, according to an agreed framework of common standards (JNCC 1998). The aim of the monitoring is to assess whether the nature conservation interest features of these sites are in favourable condition. Attributes of a particular interest feature are used to define favourable condition and targets for each attribute specify the thresholds beyond which change is of concern. English Nature has developed rapid assessment techniques to monitor the condition of over 4,000 SSSIs in England, to be supported by detailed monitoring of a small proportion of sites.

2.3 In 1998, Natural England undertook a project to investigate whether the rapid assessment method adequately indicated the condition of a grassland when compared to results obtained from more detailed information. Three National Vegetation Classification (NVC) types of lowland grasslands were examined: MG3 *Anthoxanthum odoratum-Geranium sylvaticum* grassland, CG2 *Festuca ovina-Avenula pratensis* grassland and CG5 *Bromus erectus-Brachypodium pinnatum* grassland. These NVC types were chosen to encompass a range in structure from tall meadow, through medium height pastures, to short grazed swards, as representatives of a wider variety of grassland NVC types with similar structures. The areas chosen were the Wiltshire chalk downs for CG2, the Gloucestershire Jurassic limestone of the Cotswolds for CG5 and the dales of North Yorkshire and Cumbria for MG3. Fifteen sites in total were recorded, 6 MG3 SSSIs, 6 CG2 SSSIs and 3 CG5 SSSIs. The CG sites were managed as permanent pastures and the MG3 sites as hay meadows.

2.4 One criteria for site selection was that, as far as possible, each chosen site should contain an unmixed stand of the required NVC type, and the sample plot used for detailed recording was located in this stand. Within each sample plot 40 quadrats were located in accordance with the monitoring method developed for Natural England by UCPE (Hodgson and others 1995). Each sample plot was divided into more or less equal size strips and within each strip random locations were chosen. If the random location fell on a habitat different from the NVC type under study, for example, rabbit burrow, continuous scrub (as opposed to woody individuals in a grassland sward) or trackway, the sample point was relocated. At each location a 1 x 1 m quadrat was recorded. Quadrats were divided into 9 cells, very similar to the ADAS type of nested quadrat (Critchley 1997). Nest sizes ranged from 0.004 m^2 to 1 m^2 with successive cells doubling in area (Figure 2). Recording began in the smallest cell, and the presence of all vascular plant species rooted in the cell were recorded. Presence of additional species in larger cells were recorded sequentially, moving from the smallest to the largest cell.

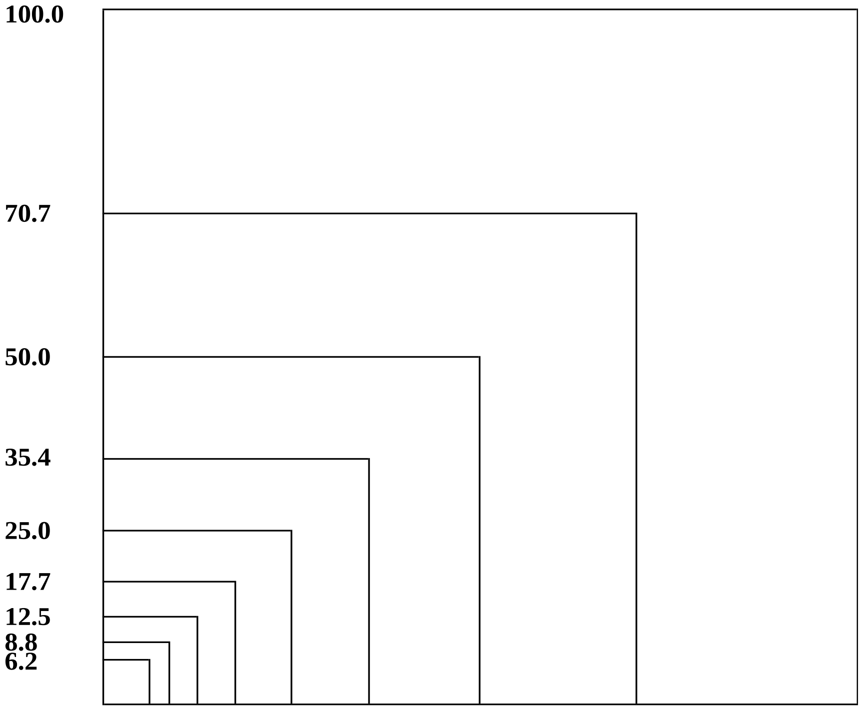


Figure 2 Dimensions of Natural England quadrat nests (cm)

Table 1 Data from Natural England Validation Study

NVC category	Site	Undamaged	Damaged
CG2	Bowerchalk Downs	40	
	Knighton Bank	40	10
	Knighton Down	40	10
	Parsonage Down	40	
	Pewsey Downs	40	
	Wylye & Church Sean Down	40	10
CG5	Cotswolds Commons	40	10
	Edge Common	40	20
	Swifts Hill	40	
MG3	Borrowbeck: Big Field	40	10
	Borrowbeck: By River	40	
	Bowber Head and Piper Hole	40	
	Muker Meadow: The Rash	40	10
	Muker Meadow: Yellands	40	10
	Pry and Bottom Meadow	40	

15 sites (6 CG2, 3 CG5, 6 MG3), 40 quadrats in each, some sites have an additional 10 or 20 quadrats categorised as damaged.

2.5 Table 1 lists the sites and data made available for the current study. In addition to the 40 quadrats per sample plot, additional quadrats were recorded at some sites in randomly selected areas designated as damaged. These damaged quadrats have not been used in the current study. A detailed description of the validation study and its findings is given in Robertson, Bingham and Slater (2000).

ADAS/Defra ESA monitoring data

2.6 When the Environmentally Sensitive Areas (ESA) scheme was launched in 1987, the Ministry of Agriculture, Fisheries and Food (MAFF), now the Department of Environment, Food and Rural Affairs (Defra), recognised a need to ensure that the scheme was delivering the desired environmental benefits. A national monitoring strategy was developed and a monitoring programme established in each ESA, covering the landscape, wildlife and historical interest. The monitoring activities are targeted and tailored according to the characteristics and environmental objectives of each ESA. The general approach has been to monitor change by establishing a baseline record of conditions when the ESA was launched and to compare this with information from subsequent surveys. Where appropriate and practicable, comparisons are also made between land that has entered the scheme (agreement land) and non-agreement land. Each ESA is reviewed by MAFF on a 5 year cycle, to assess the performance of the scheme. The national strategy for environmental monitoring of ESAs is described in the ADAS National Strategy for ESA Monitoring (ADAS, 1995).

2.7 Botanical data were collected using a field method developed by ADAS for specific use in ESA monitoring (Critchley, 1997; Critchley & Poulton, 1998). Within each monitored field, a ‘stand’ was objectively located by selecting a random distance along the diagonal between the most southerly and northerly field corners on the O.S. map. The stand was placed at least 15 m from the nearest corner to exclude the field edge zone. The four corners of the stand were marked (with galvanised metal pipes driven into the ground) for subsequent relocation with a metal detector.

2.8 Data were recorded within each of the 8 m × 4 m stands. These were divided into thirty-two 1 m × 1 m units, and species and vegetation height were recorded using nested quadrats (nests) in each of these units (see Figure 3). Plants were identified to species level where practicable. If plants could not be identified consistently in the field at this level, they were recorded to genus or as amalgams of species. Mosses and liverworts were recorded collectively, with no separation of species.

2.9 For the purpose of this study, to make the data comparable with the other datasets, information from the thirty-two 1 m^2 units was converted to a purely nested design as shown in Figure 4. Only the detailed information from the corner unit labelled 1 in Figure 3 was retained. Species presence/absence information from the other units was amalgamated into progressively larger nests and the information retained for each quadrat was the nest in which each species first occurred.

2.10 Table 2 shows the 13 ESAs from which data was used in this study. Information from an additional three ESAs was also made available but the size of the basic unit used at these sites was not 1 m^2. For simplicity these sites have been omitted from the study. Of the thirteen ESAs retained five had quadrats placed in two different “activity” areas. These activities, listed in Table 2, have been treated in this study as if they were different sites giving a total of eighteen sites in all. Some quadrats were recorded in more than one year. Where this happens the information from each year has been treated as if it were a separate quadrat.

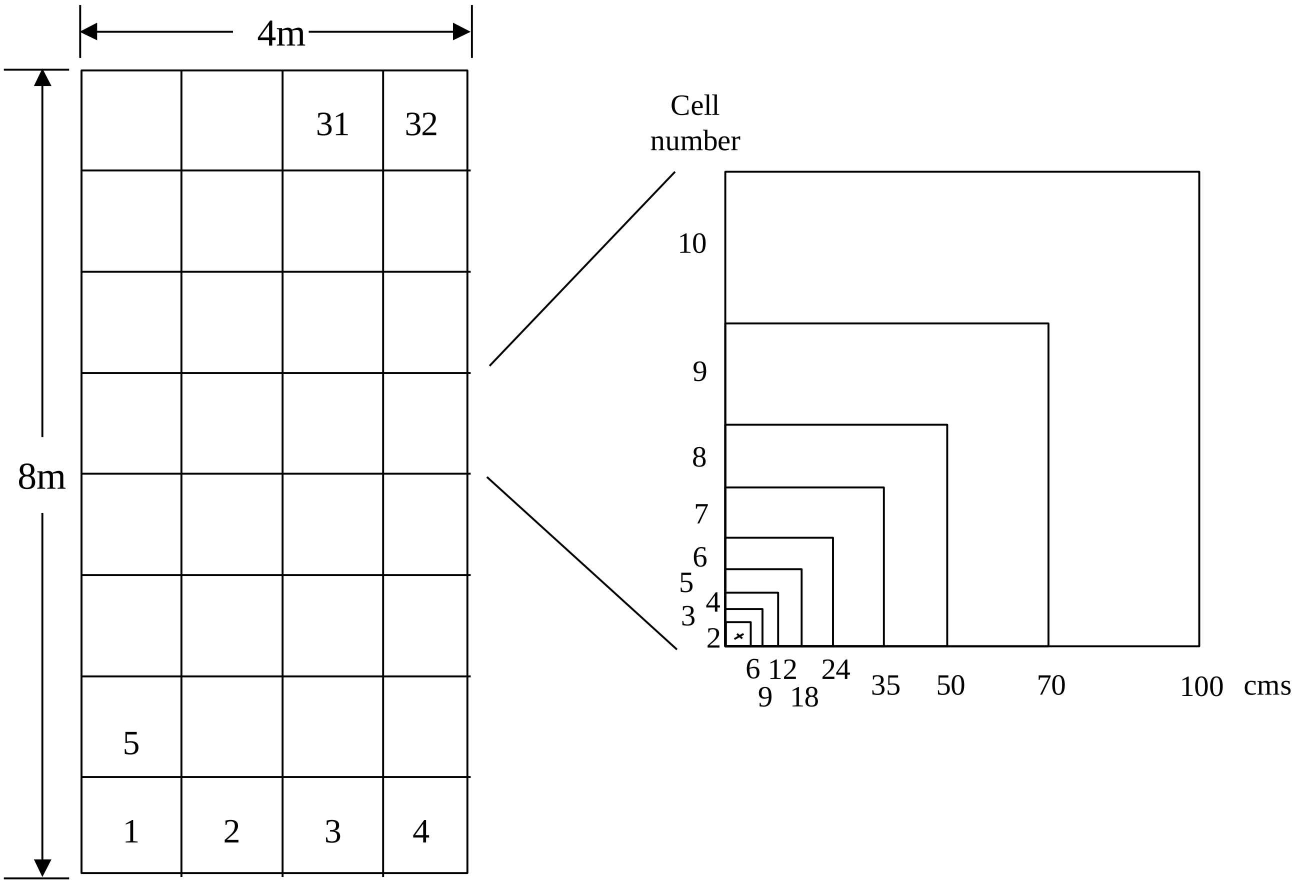
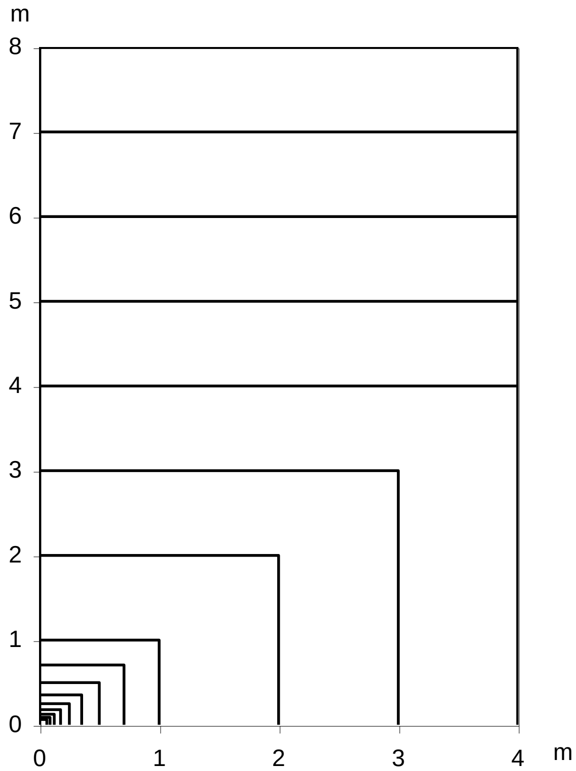


Figure 3 Sampling scheme and nest sizes for ESA monitoring



Nest sizes

cm^2	36
	81
	144
	324
	576
	1225
	2500
	4900
m^2	1
	4
	9
	16
	20
	24
	28
	32

Figure 4 Modified ESA nest sizes used in this study

Table 2 Data from ADAS ESA Study

ESA	Activity	Number of years recorded			Number of Plots
		1	2	3	
Avon Valley	Permanent grassland/downland/grazing marsh	26			26
	Wet grassland	18			18
Blackdown Hills	Permanent grassland/downland/grazing marsh	37			37
Clun ESA	Improved grassland reversion to rough grazing	15			15
	Improved grassland reversion to unimproved grassland	15			15
Dartmoor	Heather grazing and condition	31	33		64
	Permanent grassland/downland/grazing marsh	20			20
Exmoor	Grass Moorland, unimproved grassland/fell grassland		24		24
	Permanent grassland/downland/grazing marsh	25			25
Lake District	Tier 2 wetland	9	9		18
North Peak	Permanent grassland/downland/grazing marsh	35			35
Shropshire Hills	Permanent grassland/downland/grazing marsh	32			32
Somerset Levels & Moors	Wet grassland			25	25
South Wessex Downs	Permanent grassland/downland/grazing marsh	1	41		42
South West Peak	Grass Moorland, unimproved grassland/fell grassland		28		28
	Permanent grassland/downland/grazing marsh	39			39
Upper Thames Tributaries	Permanent grassland/downland/grazing marsh	40			40
West Penwith	Heathland and semi-natural grassland		28	1	29
All		**343**	**163**	**26**	**532**

13 ESA's, 18 sites, 532 plots, some plots surveyed in more than one year to give a total of 747

Countryside Survey

2.11 The Ecological Survey of Great Britain (Bunce 1984) in 1978 and the subsequent Countryside Surveys (CS) in 1990 (Barr et al. 1993) and 1998 (Haines-Young et al., 2000) were undertaken by the Natural Environment Research Council's Centre for Ecology and Hydrology (CEH), previously the Institute of Terrestrial Ecology (ITE), in order to describe the land cover, landscape features, habitats and vegetation in the wider countryside of Great Britain. Full details of survey procedures and sample locations are given in Barr et al. (1993) and Haines-Young et al. (2000). The botanical data from the three surveys, with over 13,000 reference plots in total, comprise one of the most comprehensive data sets available for quantitative analysis of vegetation change on a national scale.

2.12 The surveys were based on a stratified random sample of one kilometre squares. Vegetation sampling within each square was undertaken using a combination of randomly placed and targeted quadrats. 256 one kilometre squares were surveyed in 1978, 508 squares in 1990 and 569 squares in 1998.

2.13 Vegetation data for the 2764 randomly placed nested plots from the 1998 survey (CS2000) have been made available by CEH for the current study. These form a stratified random sample covering the whole of Great Britain. Five levels of nesting were used for these plots ranging from 4 m^2 to 200 m^2 (Figure 5). Unlike the NE and ADAS datasets all nests were centred on the same point.

2.14 During planning for CS2000 it was decided that existing tools were not appropriate for the analysis of the extensive dataset, covering the vegetation of the wider countryside at the national scale, as no pre-existing classification was designed to handle the full range of variation of the many highly disturbed situations covered. In particular the combination of the random sampling approach used and the need to detect fine changes in botanical composition over time, in conjunction with the heterogeneity of many of the samples, meant that it was not appropriate to use the existing NVC as the basis for reporting and analysis. The NVC is designed primarily to cover homogeneous stands of semi-natural vegetation. Although it includes linear features and disturbed vegetation these are rarely sufficiently homogeneous for locating relevées in. CEH therefore constructed a new classification of British vegetation, as the basic building block for the subsequent development of botanical indicators and analysis of change. This classification of vegetation in the wider countryside is known as the Countryside Vegetation System (CVS) (Bunce et al. 1999).

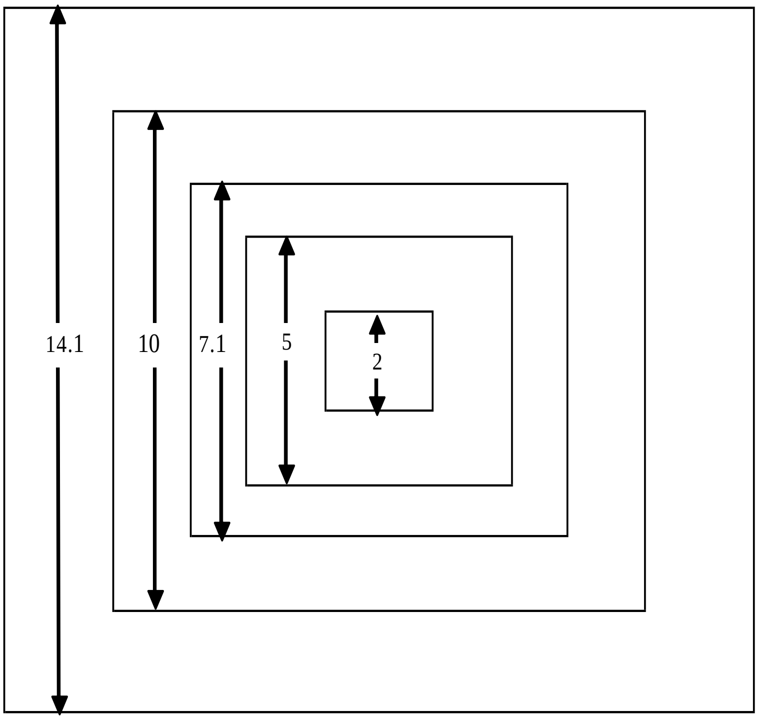


Figure 5 Nest sizes (m) used in Countryside Survey monitoring

2.15 The CVS defines 100 different baseline vegetation classes grouped into 8 aggregate classes. A ninth aggregate class encompasses quadrats not considered to have sufficient affinity to other quadrats to be placed in any of the baseline classes. This aggregate class is therefore much more heterogeneous than the others, although all CVS aggregate classes would be considered highly heterogeneous in comparison with the other datasets used in this study. Because of the random location of CS2000 quadrats there is no equivalent to the concept of a site as with the NE and ADAS datasets. The CVS aggregate classes have therefore been used in the current study as a basis for the division of plots into groups. Table 3 lists the CVS aggregate classes and shows the number of quadrats within each used in this study.

Table 3 Countryside survey data

CVS Aggregate Class		**Plots**
AG1	Crops/weeds	476
AG2	Tall grassland/herb	149
AG3	Fertile grassland	491
AG4	Infertile grassland	491
AG5	Lowland wooded	84
AG6	Upland wooded	211
AG7	Moorland grass/mosaic	350
AG8	Heath/bog	473
AG0	Other	50
All		**2764**

3 Method

3.1 Modelling frequency-area curves involves finding a mathematical expression, or formula, that can be used to predict frequency from area. Typically a general expression, that includes a small number of parameters whose values can be varied to produce different expressions of the curve, is chosen a priori. Finding a curve to represent a specific dataset then involves finding those values of the parameters that produce the best fit to the data. For example the expression

$y = abx / (a+bx)$,

where y is the species' cumulative frequency, x is quadrat size, and a and b are parameters, has been commonly used in botanical science and is used for between-species comparison in the FIBS methodology (Hodgson et al, 1995). For this curve, sometimes called the hyperbolic function, the parameters have an ecological interpretation, a is an asymptotic parameter equivalent to the maximum frequency of occurrence that the species concerned can reach, and b is the initial rate of increase of frequency with quadrat area.

3.2 The initial choice of which curve type to use is often based on previous work, as with the hyperbolic function described above. However, insufficient work has been done in this area to be sure which function provides the best representation of empirical frequency-area curves or even whether different functions may be more appropriate for different species, habitats, or at different size scales. Accordingly a range of functions was considered for use in this project. Fourteen different curves, whose formulae are listed in Table 4, were compared. These fall into two groups according to how they were derived, from either the logistic or complementary log-log functions, and the seven curves within each group vary according to the number and combination of parameters they contain. Details of the derivation of the curves are given in Appendix 1.

3.3 The maximum number of parameters is four and, like the hyperbolic curve, these have fairly simple interpretations in terms of curve shape. In each case the parameter a represents an upper limit on the frequencies that can be achieved, and b represents a lower limit. The parameter d is related to the rate of change of frequency with size and c allows for a transformation of the area scale. Hodgson et al. (1995), in choosing the hyperbolic curve argue that zero frequency is appropriate for zero quadrat size so did not therefore incorporate a lower limit to frequency other than zero into their curve choice. However it could equally well be argued that an upper limit to frequency other than one is also not appropriate, since any species will be found if quadrat size is extended to infinity. In practice the most important consideration is the fit at the limited range of sizes that are actually used in particular studies and not the behaviour of the fitted curve at the extremes of the size range.

3.4 This project used non-linear fitting techniques and maximum likelihood (see Appendix 2) to fit this range of functions to the datasets described in Section 2 in order to compare their efficiency, at the species level, in modelling frequency-area curves and in extrapolating to different scales. Details of the fitting process are given in Appendix 2. Analyses were performed using the S+ (MathSoft, 1999) and SAS (SAS Institute, 1999) statistical packages.

Table 4 Curve formulae

	Logistic	Complementary log-log
1	Four parameter curve	
	$y = a \left(\dfrac{x^c + bd}{x^c + d} \right)$	$y = a - (a - b) \exp(-dx^c)$
	The curve starts at a frequency value of *b* and rises to a maximum value of *a*. The rate of increase is determined by the parameter *d*. The parameter c produces a power transformation of the quadrat area, *x*.	
2	Three parameter curve	
	i) Frequencies start at zero for zero quadrat size	
	$y = a \left(\dfrac{x^c}{x^c + d} \right)$	$y = a(1 - \exp(-dx^c))$
	ii) Frequencies rise to 1 for very large quadrats	
	$y = \left(\dfrac{x^c + bd}{x^c + d} \right)$	$y = 1 - (1 - b) \exp(-dx^c)$
	iii) No transformation of quadrat size	
	$y = a \left(\dfrac{x + bd}{x + d} \right)$	$y = a - (a - b) \exp(-dx)$

Table continued…

	Logistic	Complementary log-log
3	Two parameter curves	
	i) Frequencies rise to 1 for very large quadrats. No transformation of quadrat size $$y = \left[\frac{x + bd}{x + d}\right]$$	$y = 1 - (1 - b) \exp(-dx)$
	ii) Frequencies start at zero. No transformation of quadrat size. Logistic curve equivalent to the hyperbolic curve used by Hodgson *et al*. (1995) $$y = a \left[\frac{x}{x + d}\right]$$	$y = a(1 - \exp(-dx))$
	iii) Frequencies start at zero and rise to 1 $$y = \left[\frac{x^c}{x^c + d}\right]$$	$y = 1 - \exp(-dx^c)$
4	One parameter curve	
	i) Frequencies start at zero and rise to 1. No transformation of quadrat size. The one parameter complementary log-log curve is an exponential curve representing complete spatial randomness $$y = \left[\frac{x}{x + d}\right]$$	$y = 1 - \exp(-dx)$

4 Frequency-area curves

4.1 This section describes the results of the analyses undertaken to determine the accuracy of a variety of mathematical functions in modelling frequency-area relationships.

4.2 An important factor in studying frequency-area relationships is whether the data under consideration represents a homogeneous stand of vegetation or is heterogeneous. Except in special circumstances it is not the case that the result obtained by averaging a set of curves with different parameters will be a curve of the same type as the individual components. It would not therefore be surprising if data from heterogeneous and homogeneous vegetation required different types of fitted curve to achieve an accurate representation.

4.3 There are two reasons why this might not be the situation for vegetation studies. Firstly, as pointed out by, for example, Hodgeson et al., (1995) and Austin (1981), it is arguable that all sites are heterogeneous to some extent and this will be particularly true for sites large enough to supply sufficient data to ensure an accurate estimate of a frequency-area curve. Secondly, if the parameters of a set of component curves are limited in the extent to which they vary then the average curve may not be too different in form anyway. Thus differences in curve type may only be apparent for very heterogeneous vegetation.

4.4 Of the main datasets considered here, the Countryside Survey dataset is the most heterogeneous since quadrats were randomly placed in vegetation. Thus even single quadrats may represent a mixture of vegetation types. The ESA monitoring data was obtained from large (32 m^2) plots placed in areas of homogeneous vegetation. However, each plot was taken from a different field. Thus individual plot data will represent homogeneous vegetation but there is likely to be a degree of heterogeneity within sets of plots. The Natural England data from lowland grassland SSSIs is the most homogeneous with forty 1 m^2 quadrats located within a relatively small homogeneous area at each site, and provision for relocating quadrats that the sampling methodology allocates to areas not considered to be the same as the NVC type under study.

Heterogeneous vegetation

4.5 As an initial step we start by considering heterogeneous data. For each of the three datasets information from all sites was combined and frequency-area curves fitted for each species. For an individual species the extent to which an accurate representation of the true underlying curve can be obtained by model fitting depends on the amount of data, that is, the number of plots in which the species is found, and the range of nest sizes across which it is first seen. It is clearly not possible, for example, to obtain a good fit for rare species that are observed in only a few plots or for species that are so common that they are found in the smallest nest size of almost all plots. For this reason model-fitting results are presented here broken down by the number of plots species occur in and by the number of different nest sizes in which they are found. In tabulating the number of nest sizes, an additional category of nest size is used to denote those plots in which the species being analysed does not occur.

4.6 Table 5 tabulates the best fitting complementary log-log model for the species found in the Natural England validation study. In total 127 species (in practise distinct names) were listed in the dataset as occurring in three or more plots. Of the eight different models considered only three provided the best fit to ten or more species, those with parameter sets *ad, cd* and *acd*. Furthermore the *ad* model only occurred as a best fit for species found in relatively few plots. In contrast the three-parameter *acd* model was increasingly found to be the best fit as the number of plots in which species were found increased. Since both the *ad* and *cd* models are special cases of the *acd* model this suggests that the most appropriate model for the majority of species in this dataset is the *acd* model, but that insufficient data can result in a less complex form being chosen. The second part of Table 5 also supports this conclusion since the *ad* model tends to be chosen when the number of nest sizes in which a species occurs is low. Conversely the *acd* model only occurs for species found close to the maximum number of nest sizes. Overall, therefore, Table 5 suggests that a species needs to be found in 20 or more plots and six or more nest sizes for an accurate choice of model to be obtained.

Table 5 Complementary log-log models for Natural England data: number of species for which specified model is best fit, by number of quadrats and number of nests in which species observed

	Parameterisation of best fitting model								
No. quadrats	**d**	**a,d**	**b,d**	**c,d**	**a,b,d**	**a,c,d**	**b,c,d**	**a,b,c,d**	**All**
3-4	3	3		2		1			**9**
5-9	2	1	1	3					**7**
10-14		2		1		1			**4**
15-19	1	2		3		1			**7**
20-29		2		5		3			**10**
30-39				2		5			**7**
40-49				3	1	4			**8**
50-99				2		11			**13**
100-199				2		29		1	**32**
200+						30			**30**
All	**6**	**10**	**1**	**23**	**1**	**85**		**1**	**127**

	Parameterisation of best fitting model								
Non-zero nests	**d**	**a,d**	**b,d**	**c,d**	**a,b,d**	**a,c,d**	**b,c,d**	**a,b,c,d**	**All**
2						1			**1**
3		1							**1**
4	4	1	1						**6**
5	1	2		2					**5**
6	1			3		1			**5**
7		3		1					**4**
8		3		2	1	2			**8**
9				6		8			**14**
10				9		73		1	**83**
All	**6**	**10**	**1**	**23**	**1**	**85**		**1**	**127**

All sites combined, model fitted to all nest sizes.

4.7 Table 6 shows the equivalent results for the seven logistic models. The same three parameter sets account for the majority of best fits and again the *acd* model is more frequently chosen as the number of plots and nest sizes in which a species occurs increases.

Table 6 Logistic models for Natural England data: number of species for which specified model is best fit, by number of quadrats and number of nests in which species observed

	Parameterisation of best fitting model								
No. quadrats	**d**	**a,d**	**b,d**	**c,d**	**a,b,d**	**a,c,d**	**b,c,d**	**a,b,c,d**	**All**
3-4	3	3		1		2			**9**
5-9	2	1	1	3					**7**
10-14		3		1					**4**
15-19	1	3		3					**7**
20-29		3		5		2			**10**
30-39		2		2	1	2			**7**
40-49		1		3	1	3			**8**
50-99		2		2	1	8			**13**
100-199		2		3	2	25			**32**
200+		4		1	2	23			**30**
All	**6**	**24**	**1**	**24**	**7**	**65**			**127**

	Parameterisation of best fitting model								
Non-zero nests	**d**	**a,d**	**b,d**	**c,d**	**a,b,d**	**a,c,d**	**b,c,d**	**a,b,c,d**	**All**
2						1			**1**
3		1							**1**
4	4		1			1			**6**
5	1	3		1					**5**
6	1	1		3					**5**
7		3		1					**4**
8		5		2	1				**8**
9		6		6	1	1			**14**
10		5		11	5	62			**83**
All	**6**	**24**	**1**	**24**	**7**	**65**			**127**

All sites combined, model fitted to all nest sizes.

4.8 Table 7 tabulates the complementary log-log model fitting results for the CS data. Models were fitted to the 544 species found in more than three plots. For this dataset the *cd* model was the best fitting model for almost fifty percent of species. The other two and one parameter models were only chosen for species found in few plots suggesting that their choice was a result of insufficient data. Three and four parameter models are chosen to a lesser extent than the *cd* model but tend to occur more often for species found in many plots and/or the majority of nest sizes.

Table 7 Complementary log-log models for Countryside Survey data: number of species for which specified model is best fit, by number of quadrats and number of nests in which species observed

	Parameterisation of best fitting model								
No. quadrats	**d**	**a,d**	**b,d**	**c,d**	**a,b,d**	**a,c,d**	**b,c,d**	**a,b,c,d**	**All**
3-4	15	18	24	29		5	1	5	**97**
5-9	2	20	35	56	4		1	6	**124**
10-14		3	10	33	3	1		5	**55**
15-19			12	19	3	2			**36**
20-29			1	27	5	1		2	**36**
30-39			1	21	3			1	**26**
40-49				8	6		2		**16**
50-99			2	31	12		8	1	**54**
100-199				19	17	1	8	1	**46**
200+				25	13		9	7	**54**
All	**17**	**41**	**85**	**268**	**66**	**10**	**29**	**28**	**544**

	Parameterisation of best fitting model								
Non-zero nests	**d**	**a,d**	**b,d**	**c,d**	**a,b,d**	**a,c,d**	**b,c,d**	**a,b,c,d**	**All**
2				13		2			**15**
3	10	21	13	7		3	2	4	**60**
4	6	10	33	41	3	1		11	**105**
5	1	7	27	30	6	1		3	**75**
6		3	12	177	57	3	27	10	**289**
All	**17**	**41**	**85**	**268**	**66**	**10**	**29**	**28**	**544**

All vegetation classes combined, model fitted to all nest sizes.

4.9 Table 8 shows the results of fitting the seven logistic models to the CS data. The results are broadly similar to those obtained for the complementary log-log models. The *cd* model is by far the most often chosen model. Models with lower numbers of parameters are only chosen for species occurring in lower numbers of plots and those with higher numbers of parameters only for species occurring in higher numbers of plots and the maximum number of nests.

Table 8 Logistic models for Countryside Survey data**:** number of species for which specified model is best fit, by number of quadrats and number of nests in which species observed

	Parameterisation of best fitting model								
No. quadrats	**d**	**a,d**	**b,d**	**c,d**	**a,b,d**	**a,c,d**	**b,c,d**	**a,b,c,d**	**All**
3-4	14	26	24	22		5	1	5	**97**
5-9	2	29	41	46		1	1	4	**124**
10-14		8	12	30	1			4	**55**
15-19		3	12	19	1			1	**36**
20-29		5	1	26	2			2	**36**
30-39			1	21	2			2	**26**
40-49				8	5		2	1	**16**
50-99			2	31	13		8		**54**
100-199				15	23	1	5	2	**46**
200+				18	28		7	1	**54**
All	**16**	**71**	**93**	**236**	**75**	**7**	**24**	**22**	**544**

	Parameterisation of best fitting model								
Non-zero nests	**d**	**a,d**	**b,d**	**c,d**	**a,b,d**	**a,c,d**	**b,c,d**	**a,b,c,d**	**All**
2		9	4			2			**15**
3	10	21	14	7		3	2	3	**60**
4	5	19	35	37	1			8	**105**
5	1	13	28	28		1		4	**75**
6		9	12	164	74	1	22	7	**289**
All	**16**	**71**	**93**	**236**	**75**	**7**	**24**	**22**	**544**

All vegetation classes combined, model fitted to all nest sizes.

4.10 Tables 9 and 10 show the model fitting results for the ADAS dataset. These are even more clear-cut than the other two datasets, very similar for both types of model, and substantially the same as for the Natural England dataset. Two parameterisations account for three-quarters of the chosen models, *cd* and *acd*. Instances in which other models are chosen are largely confined to species found in fewer than twenty plots or less than 10 nest sizes.

4.11 The range of nest sizes used in recording the ADAS dataset is almost as great as that from the CS and Natural England datasets combined while the latter two datasets have no overlap in nest sizes. In order to see whether the apparent differences in results described above are related to the range of nest sizes, two additional datasets were generated from the ADAS data. The first contained only information from the seven nest sizes greater than 1 m^2 and was thus comparable to the CS dataset. The second contained only information from the nine nest sizes $1m^2$ or smaller and was therefore comparable with the NE dataset.

Table 9 Complementary log-log models for ADAS ESA data: number of species for which specified model is best fit, by number of quadrats and number of nests in which species observed

	Parameterisation of best fitting model								
No. quadrats	**d**	**a,d**	**b,d**	**c,d**	**a,b,d**	**a,c,d**	**b,c,d**	**a,b,c,d**	**All**
3-4	13	7	1	14	2	5			**42**
5-9	14	11	1	21		5			**52**
10-14	1	6	1	16	1	4			**29**
15-19		4	1	12	4	6			**27**
20-29		1		16		6			**23**
30-39		1		9		9			**19**
40-49				13		5	1		**19**
50-99				26	1	31		1	**59**
100-199				11		21	1		**33**
200+				1		23			**24**
All	**28**	**30**	**4**	**139**	**8**	**115**	**2**	**1**	**327**

	Parameterisation of best fitting model								
Non-zero nests	**d**	**a,d**	**b,d**	**c,d**	**a,b,d**	**a,c,d**	**b,c,d**	**a,b,c,d**	**All**
3	2	3		1		5			**11**
4	11	9		10	2	1			**33**
5	7	1	2	10		1			**21**
6	5	6		7	1	3			**22**
7	2	6	1	9	1	3			**22**
8	1	3		10	2				**16**
9		2	1	6		5			**14**
10				8	1	2			**11**
11				7		8			**15**
12				13		5			**18**
13				8	1	7		1	**17**
14				8		11	1		**20**
15				12		12			**24**
16				16		14			**30**
17				14		38	1		**53**
All	**28**	**30**	**4**	**139**	**8**	**115**	**2**	**1**	**327**

All sites combined, model fitted to all nest sizes.

Table 10 Logistic models for ADAS ESA data: number of species for which specified model is best fit, by number of quadrats and number of nests in which species observed

	Parameterisation of best fitting model								
No. quadrats	**d**	**a,d**	**b,d**	**c,d**	**a,b,d**	**a,c,d**	**b,c,d**	**a,b,c,d**	**All**
3-4	13	7	1	13	2	6			**42**
5-9	14	13	1	19	1	4			**52**
10-14	1	9	1	14	2	2			**29**
15-19		4	2	12	1	6		2	**27**
20-29		1		16		6			**23**
30-39		2		9		8			**19**
40-49				12		5	2		**19**
50-99				26	2	31			**59**
100-199				12		20	1		**33**
200+				4		20			**24**
All	**28**	**36**	**5**	**137**	**8**	**108**	**3**	**2**	**327**

	Parameterisation of best fitting model								
Non-zero nests	**d**	**a,d**	**b,d**	**c,d**	**a,b,d**	**a,c,d**	**b,c,d**	**a,b,c,d**	**All**
3	2	2				7			**11**
4	11	7		9	3	3			**33**
5	7	3	2	9					**21**
6	5	8		7		1		1	**22**
7	2	8	1	8	1	1		1	**22**
8	1	3		10	2				**16**
9		4	1	6		3			**14**
10			1	7		3			**11**
11		1		7		7			**15**
12				13		5			**18**
13				7	2	8			**17**
14				8		11	1		**20**
15				13		11			**24**
16				15		14	1		**30**
17				18		34	1		**53**
All	**28**	**36**	**5**	**137**	**8**	**108**	**3**	**2**	**327**

All sites combined, model fitted to all nest sizes.

4.12 Tables 11 and 12 show the results for the two types of model fitted to the large ADAS nest sizes. The results are surprisingly similar to the CS results. The *cd* model is most often chosen, other two or one parameter models are chosen mainly for species found in fewer plots, and the *acd* model, almost as popular as the *cd* model in the full dataset, is rarely chosen as the best fit.

Table 11 Complementary log-log models for ADAS ESA data: number of species for which specified model is best fit, by number of quadrats and number of nests in which species observed

	Parameterisation of best fitting model								
No. quadrats	**d**	**a,d**	**b,d**	**c,d**	**a,b,d**	**a,c,d**	**b,c,d**	**a,b,c,d**	**All**
3-4	17	13	5	5		1	1		**42**
5-9	13	15	6	10		1	2	5	**52**
10-14	1	13	9	5			1		**29**
15-19	1	10	2	12		1		1	**27**
20-29		4	4	14				1	**23**
30-39		4	1	12				2	**19**
40-49		1	5	9	1	1		2	**19**
50-99		2	8	39	2	8			**59**
100-199			1	26	4	1		1	**33**
200+				12	7	4		1	**24**
All	**32**	**62**	**41**	**144**	**14**	**17**	**4**	**13**	**327**

	Parameterisation of best fitting model								
Non-zero nests	**d**	**a,d**	**b,d**	**c,d**	**a,b,d**	**a,c,d**	**b,c,d**	**a,b,c,d**	**All**
2		5		2					**7**
3	2	13	5	4		2	1	2	**29**
4	14	17	5	11	1		2	5	**55**
5	7	15	7	12	1	1	1	2	**46**
6	5	7	6	22	2	5			**47**
7	3	3	10	39		5		4	**64**
8	1	2	8	54	10	4			**79**
All	**32**	**62**	**41**	**144**	**14**	**17**	**4**	**13**	**327**

All sites combined, model fitted to nest sizes above 1 m^2.

Table 12 Logistic models for ADAS ESA data: number of species for which specified model is best fit, by number of quadrats and number of nests in which species observed

	Parameterisation of best fitting model								
No. quadrats	**d**	**a,d**	**b,d**	**c,d**	**a,b,d**	**a,c,d**	**b,c,d**	**a,b,c,d**	**All**
3-4	17	6	6	5		6	1	1	**42**
5-9	15	15	8	5		2	2	5	**52**
10-14	1	9	10	5		3	1		**29**
15-19	1	10	2	11		2		1	**27**
20-29		9	4	9				1	**23**
30-39		9	2	7				1	**19**
40-49		4	5	8				2	**19**
50-99		17	9	32		1			**59**
100-199		3	3	25	1			1	**33**
200+				14	5	4		1	**24**
All	**34**	**82**	**49**	**121**	**6**	**18**	**4**	**13**	**327**

	Parameterisation of best fitting model								
Non-zero nests	**d**	**a,d**	**b,d**	**c,d**	**a,b,d**	**a,c,d**	**b,c,d**	**a,b,c,d**	**All**
2		3	4						**7**
3	2	5	6	4		8	1	3	**29**
4	14	16	6	8		3	2	6	**55**
5	9	21	7	6		2	1		**46**
6	5	20	6	15		1			**47**
7	3	12	10	34		1		4	**64**
8	1	5	10	54	6	3			**79**
All	**34**	**82**	**49**	**121**	**6**	**18**	**4**	**13**	**327**

All sites combined, model fitted to nest sizes above 1 m^2.

4.13 Tables 13 and 14 show results for the smaller ADAS nest sizes. These are now close to the results for the NE data. The patterns of occurrence are very similar and the same three parameterisations, *ad, cd* and *acd*, account for the majority of best fitting model.

Table 13 Complementary log-log models for ADAS ESA data: number of species for which specified model is best fit, by number of quadrats and number of nests in which species observed

	Parameterisation of best fitting model								
No. quadrats	**d**	**a,d**	**b,d**	**c,d**	**a,b,d**	**a,c,d**	**b,c,d**	**a,b,c,d**	**All**
3-4	11	10	1	9		1	2		**34**
5-9		14	2	22		1			**39**
10-14		6	1	9	3	7			**26**
15-19				8		6			**14**
20-29		3		10	1	7			**21**
30-39				7	1	11		1	**20**
40-49				4	1	13			**18**
50-99		1		8		26			**35**
100-199						17			**17**
200+						10			**10**
All	**11**	**34**	**4**	**77**	**6**	**99**	**2**	**1**	**234**

	Parameterisation of best fitting model								
Non-zero nests	**d**	**a,d**	**b,d**	**c,d**	**a,b,d**	**a,c,d**	**b,c,d**	**a,b,c,d**	**All**
2				2		1			**3**
3	3	2		3			2		**10**
4	7	8	2	5					**22**
5	1	7	1	6					**15**
6		6	1	11	2	4			**24**
7		6		9	1	2			**18**
8		2		9	1	7		1	**20**
9		3		11	2	18			**34**
10				21		67			**88**
All	**11**	**34**	**4**	**77**	**6**	**99**	**2**	**1**	**234**

All sites combined, model fitted to all nest sizes 1 m^2 and below.

Table 14 Logistic models for ADAS ESA data: number of species for which specified model is best fit, by number of quadrats and number of nests in which species observed

	Parameterisation of best fitting model								
No. quadrats	**d**	**a,d**	**b,d**	**c,d**	**a,b,d**	**a,c,d**	**b,c,d**	**a,b,c,d**	**All**
3-4	9	13	4	5		3			**34**
5-9		16	2	20				1	**39**
10-14		12	1	9	2	1		1	**26**
15-19		6		6		2			**14**
20-29		5		10	2	4			**21**
30-39		1		7	3	8		1	**20**
40-49		3		4	2	9			**18**
50-99		4		7	2	22			**35**
100-199						17			**17**
200+						10			**10**
All	**9**	**60**	**7**	**68**	**11**	**76**		**3**	**234**

	Parameterisation of best fitting model								
Non-zero nests	**d**	**a,d**	**b,d**	**c,d**	**a,b,d**	**a,c,d**	**b,c,d**	**a,b,c,d**	**All**
2				2		1			**3**
3	3	2		3			2		**10**
4	7	8	2	5					**22**
5	1	7	1	6					**15**
6		6	1	11	2	4			**24**
7		6		9	1	2			**18**
8		2		9	1	7		1	**20**
9		3		11	2	18			**34**
10				21		67			**88**
All	**11**	**34**	**4**	**77**	**6**	**99**	**2**	**1**	**234**

All sites combined, model fitted to all nest sizes 1 m^2 and below.

Homogeneous data

4.14 We turn now to more homogeneous data. For the NE and ADAS datasets models were fitted to species from each of the sites separately. For the CS datasets the CVS aggregate classes were treated as if they were separate sites. As discussed above the degree of homogeneity will vary across the datasets with the NE sites being most homogeneous and the CS classes least.

4.15 Tables 15 to 24 are the site-specific equivalents of Tables 5 to 14. In examining these tables it should be borne in mind that examining species at the site level has a number of implications that will impact on the results. Firstly the number of plots in which most species are found will be substantially reduced. As a consequence fitted models are likely to be less accurate and there will be more variability in the choice of the best fitting model. Secondly many species no longer occur sufficiently often to make model fitting feasible at all. For example, 544 species were modelled for the CS data when all CVS classes were combined but this only increased to 1245 when the nine aggregate classes were modelled separately. This is partially due to the occurrence of species that are highly class specific but more to the reduction in plot numbers for less common species. The consequence is that a different set of species is being examined than in the previous tables. Thirdly the categories used for the number of plots in which species occur do not have quiet the same interpretation. For example, a species which occurred in 40 plots when all the NE sites were combined would be present in only 7% of plots, a relatively uncommon species, whereas a species occurring in 40 plots at the site level would be present in 100% of plots.

4.16 Taking these points into consideration, however, the results from fitting frequency-area curves to site specific data are similar to those from the all sites data. Tables 15 and 16 show the best fitting models for the NE data. As before the *ad, cd* and *acd* models are chosen more often as best fit. The *cd* parameterisation is chosen most frequently for both complementary log-log and logistic types with *acd* second choice of the complementary log-log models and *ad* second for the logistic.

Table 15 Complementary log-log models for Natural England data: number of species for which specified model is best fit, by number of quadrats and number of nests in which species observed

	Parameterisation of best fitting model								
No. quadrats	**d**	**a,d**	**b,d**	**c,d**	**a,b,d**	**a,c,d**	**b,c,d**	**a,b,c,d**	**All**
3-4	19	18	5	24		5		1	**72**
5-9	21	27	4	41	1	5	1		**100**
10-14	3	19	2	40	2	13			**79**
15-19		9	1	30	5	14			**59**
20-29	1	7	2	62	1	40	1		**114**
30-39		2		69	12	114			**197**
40-49			9	63	1				**73**
All	**44**	**82**	**23**	**329**	**22**	**191**	**2**	**1**	**694**

	Parameterisation of best fitting model								
Non-zero nests	**d**	**a,d**	**b,d**	**c,d**	**a,b,d**	**a,c,d**	**b,c,d**	**a,b,c,d**	**All**
1			2						**2**
2				8		1			**9**
3	10	5		9		4			**28**
4	17	10	4	28	1	2	1	1	**64**
5	12	13	4	22	2	9			**62**
6	5	23	7	39	5	12	1		**92**
7		15	2	42	2	17			**78**
8		8	3	39	7	33			**90**
9		8	1	67	3	72			**151**
10				75	2	41			**118**
All	**44**	**82**	**23**	**329**	**22**	**191**	**2**	**1**	**694**

Sites analysed separately, model fitted to all nest sizes.

Table 16 Logistic models for Natural England data: number of species for which specified model is best fit, by number of quadrats and number of nests in which species observed

	Parameterisation of best fitting model								
No. quadrats	**d**	**a,d**	**b,d**	**c,d**	**a,b,d**	**a,c,d**	**b,c,d**	**a,b,c,d**	**All**
3-4	19	20	6	16		9		2	**72**
5-9	25	36	4	31		3	1		**100**
10-14	7	34	2	33	2	1			**79**
15-19	2	25	1	27	2	2			**59**
20-29	8	36	3	57	3	5	2		**114**
30-39	27	65	21	69	4	6		5	**197**
40-49	37		2	20			14		**73**
All	**125**	**216**	**39**	**253**	**11**	**26**	**17**	**7**	**694**

	Parameterisation of best fitting model								
Non-zero nests	**d**	**a,d**	**b,d**	**c,d**	**a,b,d**	**a,c,d**	**b,c,d**	**a,b,c,d**	**All**
1	2								**2**
2	2			6		1			**9**
3	13	4		2		6	2	1	**28**
4	21	14	4	17		4	3	1	**64**
5	19	16	5	14		5	1	2	**62**
6	22	34	5	22		2	7		**92**
7	11	35		29	1		2		**78**
8	15	32	8	27	3	2	1	2	**90**
9	14	57	7	63	4	4	1	1	**151**
10	6	24	10	73	3	2			**118**
All	**125**	**216**	**39**	**253**	**11**	**26**	**17**	**7**	**694**

Sites analysed separately, model fitted to all nest sizes.

4.17 Tables 17 and 18 show results for the CS data. Again the *cd* parameterisation is most often chosen for both curve types with other two and one parameter curves mainly being chosen for species that occur in small numbers of plots (approximately <20). Models with three or four parameters are chosen less frequently but there is a suggestion that they are chosen more often for species that occur in the highest number of nests.

Table 17 Complementary log-log models for Countryside Survey data: number of species for which specified model is best fit, by number of quadrats and number of nests in which species observed

	Parameterisation of best fitting model								
No. quadrats	**d**	**a,d**	**b,d**	**c,d**	**a,b,d**	**a,c,d**	**b,c,d**	**a,b,c,d**	**All**
3-4	49	60	67	89	1	18	9	15	**308**
5-9	15	61	60	159	10	1	6	20	**332**
10-14	2	7	41	82	9	2	1	4	**148**
15-19		3	12	40	9	2		2	**68**
20-29		1	7	68	16	4		6	**102**
30-39			4	37	12		1	1	**55**
40-49				23	11		5		**39**
50-99			2	51	27		8	2	**90**
100-199				30	19	1	7		**57**
200+				23	15	1	7		**46**
All	**66**	**132**	**193**	**602**	**129**	**29**	**44**	**50**	**1245**

	Parameterisation of best fitting model								
Non-zero nests	**d**	**a,d**	**b,d**	**c,d**	**a,b,d**	**a,c,d**	**b,c,d**	**a,b,c,d**	**All**
2		1		35	6	3			**45**
3	17	54	36	21		14	15	17	**174**
4	37	37	73	103	6	5	1	19	**281**
5	11	31	53	119	13	1		11	**239**
6	1	9	31	324	104	6	28	3	**506**
All	**66**	**132**	**193**	**602**	**129**	**29**	**44**	**50**	**1245**

Vegetation classes analysed separately, model fitted to all nest sizes.

Table 18 Logistic models for Countryside Survey data: number of species for which specified model is best fit, by number of quadrats and number of nests in which species observed

	Parameterisation of best fitting model								
No. quadrats	**d**	**a,d**	**b,d**	**c,d**	**a,b,d**	**a,c,d**	**b,c,d**	**a,b,c,d**	**All**
3-4	51	58	91	69		19	10	10	**308**
5-9	15	86	74	127		3	7	20	**332**
10-14	2	19	40	80			1	6	**148**
15-19		9	12	42				5	**68**
20-29		5	9	73	5			10	**102**
30-39			4	39	10		1	1	**55**
40-49			1	24	5		4	5	**39**
50-99		1	2	53	21		6	7	**90**
100-199				24	22	1	8	2	**57**
200+				17	20		6	3	**46**
All	**68**	**178**	**233**	**548**	**83**	**23**	**43**	**69**	**1245**

	Parameterisation of best fitting model								
Non-zero nests	**d**	**a,d**	**b,d**	**c,d**	**a,b,d**	**a,c,d**	**b,c,d**	**a,b,c,d**	**All**
2		5	35	2		3			**45**
3	20	48	40	23		17	17	9	**174**
4	36	60	73	88		2	1	21	**281**
5	11	48	52	113	1			14	**239**
6	1	17	33	322	82	1	25	25	**506**
All	**68**	**178**	**233**	**548**	**83**	**23**	**43**	**69**	**1245**

Vegetation classes analysed separately, model fitted to all nest sizes.

4.18 Tables 19 and 20 give results for the complete ADAS dataset. Again the *cd* parameterisation is overwhelmingly the most common with the *acd* parameterisation also accounting for a substantial proportion of species. The *d, ad* and *bd* parameterisations are mainly chosen only for species occurring in less that 20 plots or fewer than six nest sizes. As before results from the ADAS data for the larger nest sizes (Tables 21 and 22) are similar to the CS data and those from the smaller nest sizes (Tables 23 and 24) are similar to the NE results.

Table 19 Complementary log-log models for ADAS ESA data: number of species for which specified model is best fit, by number of quadrats and number of nests in which species observed

	Parameterisation of best fitting model								
No. quadrats	**d**	**a,d**	**b,d**	**c,d**	**a,b,d**	**a,c,d**	**b,c,d**	**a,b,c,d**	**All**
3-4	76	73	10	81	8	36	1	1	**286**
5-9	46	67	6	171	6	24	3	3	**326**
10-14	2	11	2	112	4	25	6		**162**
15-19	3	6		72	2	26	4		**113**
20-29		2		81		35	4	1	**123**
30-39				51		24	2		**77**
40-49				23		11	5		**39**
50-99				44		30		1	**75**
All	**127**	**159**	**18**	**635**	**20**	**211**	**25**	**6**	**1201**

	Parameterisation of best fitting model								
Non-zero nests	**d**	**a,d**	**b,d**	**c,d**	**a,b,d**	**a,c,d**	**b,c,d**	**a,b,c,d**	**All**
2		1		2		4			**7**
3	20	23	2	20	1	13	1	1	**81**
4	57	67	6	51	7	17	1	2	**208**
5	23	27	5	58	6	10		1	**130**
6	14	27	3	72	3	9	1		**129**
7	11	9	1	69	2	18	5		**115**
8	2	3	1	65	1	12	4	1	**89**
9		2		50		20	1		**73**
10				52		18	6		**76**
11				55		21	1		**77**
12				39		22	2	1	**64**
13				28		12	1		**41**
14				30		15	1		**46**
15				24		17	1		**42**
16				16		3			**19**
17				4					**4**
All	**127**	**159**	**18**	**635**	**20**	**211**	**25**	**6**	**1201**

Sites analysed separately, model fitted to all nest sizes.

Table 20 Logistic models for ADAS ESA data: number of species for which specified model is best fit, by number of quadrats and number of nests in which species observed

	Parameterisation of best fitting model								
No. quadrats	**d**	**a,d**	**b,d**	**c,d**	**a,b,d**	**a,c,d**	**b,c,d**	**a,b,c,d**	**All**
3-4	78	88	15	68	3	29	1	4	**286**
5-9	53	80	7	159	9	11	2	5	**326**
10-14	3	23	3	108	6	14	4	1	**162**
15-19	12	9	1	70	5	11	4	1	**113**
20-29	3	5	2	88	2	18	5		**123**
30-39	2	1	1	55		14	4		**77**
40-49				21		11	7		**39**
50-99	1			57	1	12	4		**75**
All	**152**	**206**	**29**	**626**	**26**	**120**	**31**	**11**	**1201**

	Parameterisation of best fitting model								
Non-zero nests	**d**	**a,d**	**b,d**	**c,d**	**a,b,d**	**a,c,d**	**b,c,d**	**a,b,c,d**	**All**
2			2			5			**7**
3	23	20	3	11		21	1	2	**81**
4	60	75	8	44	5	10	1	5	**208**
5	25	38	6	55	3	1		2	**130**
6	21	34	3	66	3	1		1	**129**
7	15	24	3	60	5	4	3	1	**115**
8	8	5	3	62	5	4	2		**89**
9		8		51	1	12	1		**73**
10		2	1	54	2	10	7		**76**
11				60	1	12	4		**77**
12				45	1	15	3		**64**
13				31		7	3		**41**
14				39		6	1		**46**
15				28		9	5		**42**
16				16		3			**19**
17				4					**4**
All	**152**	**206**	**29**	**626**	**26**	**120**	**31**	**11**	**1201**

Sites analysed separately, model fitted to all nest sizes.

Table 21 Complementary log-log models for ADAS ESA data: number of species for which specified model is best fit, by number of quadrats and number of nests in which species observed

	Parameterisation of best fitting model								
No. quadrats	**d**	**a,d**	**b,d**	**c,d**	**a,b,d**	**a,c,d**	**b,c,d**	**a,b,c,d**	**All**
3-4	85	89	33	43		17	14	5	**286**
5-9	49	95	69	71		9	9	24	**326**
10-14	6	43	53	46		2	3	9	**162**
15-19	14	25	30	35			2	7	**113**
20-29	6	22	26	61		1	3	4	**123**
30-39	4	10	16	39	1	3		4	**77**
40-49	2	4	7	21	2		1	2	**39**
50-99	4	2	18	38	6	5	1	1	**75**
All	**170**	**290**	**252**	**354**	**9**	**37**	**33**	**56**	**1201**

	Parameterisation of best fitting model								
Non-zero nests	**d**	**a,d**	**b,d**	**c,d**	**a,b,d**	**a,c,d**	**b,c,d**	**a,b,c,d**	**All**
1	16								**16**
2	12	33	7	37		4	1		**94**
3	21	82	52	33		18	18	16	**240**
4	62	112	58	79	1	5	7	29	**353**
5	30	46	49	85	5	6	3	4	**228**
6	21	15	46	75	2	4	2	5	**170**
7	7	2	30	37	1		2	2	**81**
8	1		10	8					**19**
All	**170**	**290**	**252**	**354**	**9**	**37**	**33**	**56**	**1201**

Sites analysed separately, model fitted to nest sizes above 1 m^2.

Table 22 Logistic models for ADAS ESA data: number of species for which specified model is best fit, by number of quadrats and number of nests in which species observed

	Parameterisation of best fitting model								
No. quadrats	**d**	**a,d**	**b,d**	**c,d**	**a,b,d**	**a,c,d**	**b,c,d**	**a,b,c,d**	**All**
3-4	78	53	66	26		34	14	15	**286**
5-9	61	85	77	44		21	9	29	**326**
10-14	18	39	57	26		9	6	7	**162**
15-19	24	23	36	19		3	3	5	**113**
20-29	13	45	27	28		1	5	4	**123**
30-39	10	24	19	18		1	1	4	**77**
40-49	2	16	7	10			2	2	**39**
50-99	12	18	21	21		1	1	1	**75**
All	**218**	**303**	**310**	**192**		**70**	**41**	**67**	**1201**

	Parameterisation of best fitting model								
Non-zero nests	**d**	**a,d**	**b,d**	**c,d**	**a,b,d**	**a,c,d**	**b,c,d**	**a,b,c,d**	**All**
1	16								**16**
2	11	6	69	3		4	1		**94**
3	20	46	46	33		50	21	24	**240**
4	72	134	57	35		13	8	34	**353**
5	45	74	46	51		3	6	3	**228**
6	38	34	46	44			3	5	**170**
7	15	7	34	22			2	1	**81**
8	1	2	12	4					**19**
All	**218**	**303**	**310**	**192**		**70**	**41**	**67**	**1201**

Sites analysed separately, model fitted to nest sizes above 1 m^2.

Table 23 Complementary log-log models for ADAS ESA data: number of species for which specified model is best fit, by number of quadrats and number of nests in which species observed

	Parameterisation of best fitting model								
No. quadrats	**d**	**a,d**	**b,d**	**c,d**	**a,b,d**	**a,c,d**	**b,c,d**	**a,b,c,d**	**All**
3-4	34	60	7	59	1	14	4	5	**184**
5-9	10	44	9	92	4	17		5	**181**
10-14		8	3	54	11	15			**91**
15-19		2	2	39	4	28		1	**76**
20-29		1		38	9	26		1	**75**
30-39				27	2	19		1	**49**
40-49		1		8	1	9			**19**
50-99				9		18			**27**
All	**44**	**116**	**21**	**326**	**32**	**146**	**4**	**13**	**702**

	Parameterisation of best fitting model								
Non-zero nests	**d**	**a,d**	**b,d**	**c,d**	**a,b,d**	**a,c,d**	**b,c,d**	**a,b,c,d**	**All**
2				4					**4**
3	14	9	3	13		11	4	2	**56**
4	20	49	6	41	1	4		6	**127**
5	7	31	7	35	5	9		2	**96**
6	3	13	4	52	9	17		1	**99**
7		10	1	45	7	14		1	**78**
8		2		49	3	22		1	**77**
9		1		51	7	31			**90**
10		1		36		38			**75**
All	**44**	**116**	**21**	**326**	**32**	**146**	**4**	**13**	**702**

Sites analysed separately, model fitted to nest sizes 1 m^2 and below.

Table 24 Logistic models for ADAS ESA data: number of species for which specified model is best fit, by number of quadrats and number of nests in which species observed

	Parameterisation of best fitting model								
No. quadrats	**d**	**a,d**	**b,d**	**c,d**	**a,b,d**	**a,c,d**	**b,c,d**	**a,b,c,d**	**All**
3-4	34	60	7	59	1	14	4	5	**184**
5-9	10	44	9	92	4	17		5	**181**
10-14		8	3	54	11	15			**91**
15-19		2	2	39	4	28		1	**76**
20-29		1		38	9	26		1	**75**
30-39				27	2	19		1	**49**
40-49		1		8	1	9			**19**
50-99				9		18			**27**
All	**44**	**116**	**21**	**326**	**32**	**146**	**4**	**13**	**702**

	Parameterisation of best fitting model								
Non-zero nests	**d**	**a,d**	**b,d**	**c,d**	**a,b,d**	**a,c,d**	**b,c,d**	**a,b,c,d**	**All**
2			2	2					**4**
3	16	11	5	6		10	2	6	**56**
4	19	50	8	30		12		8	**127**
5	7	43	8	30	1	2		5	**96**
6	8	35	5	42	5	2		2	**99**
7	2	30	5	35	3	1		2	**78**
8	2	22	2	42	3	4	1	1	**77**
9		17	1	51	10	11			**90**
10		12		39	2	22			**75**
All	**54**	**220**	**36**	**277**	**24**	**64**	**3**	**24**	**702**

Sites analysed separately, model fitted to nest sizes 1 m^2 and below.

Model type

4.19 So far the two types of curve have been considered separately. Tables 25 to 29 tabulate the best fitting models when both curve types are compared. Results are given for three categories of species, those found in 3 or more plots, in 10 or more plots and in 20 or more plots, in order to show the effect on model selection of increasing amounts of data. For the NE data (Table 25) it is clear that the logistic curve provides the best fit to the data whether sites are combined or not. In contrast the best fitting models are approximately equally divided between the two curve types for the CS data (Table 26). The ADAS data (Table 27) is more like the NE data in that the logistic curve is most often the best fitting model. This is also true for the ADAS dataset restricted to small nests (Table 28) whereas restriction to large nests (Table 29) gives a pattern more like the CS dataset though the logistic curve still performs slightly better.

Table 25a Natural England - comparison of model types: All sites combined

	Best fitting model								
	d	**a,d**	**b,d**	**c,d**	**a,b,d**	**a,c,d**	**b,c,d**	**a,b,c,d**	**All**
Species occurring in 3+ plots									
complementary log-log	1	6	1	10		4			**22**
logistic	5	19		14	6	61			**105**
All	**6**	**25**	**1**	**24**	**6**	**65**			**127**
Species occurring in 10+ plots									
complementary log-log		3		6		3			**12**
logistic	1	17		14	6	61			**99**
All	**1**	**20**		**20**	**6**	**64**			**111**
Species occurring in 20+ plots									
complementary log-log				4		3			**7**
logistic		14		12	6	61			**93**
All		**14**		**16**	**6**	**64**			**100**

Table 25b Natural England - comparison of model types: Sites analysed separately

	Best fitting model								
	d	**a,d**	**b,d**	**c,d**	**a,b,d**	**a,c,d**	**b,c,d**	**a,b,c,d**	**All**
Species occurring in 3+ plots									
complementary log-log	12	42	14	83	12	18			**181**
logistic	99	176	22	180	8	16	8	4	**513**
All	**111**	**218**	**36**	**263**	**20**	**34**	**8**	**4**	**694**
Species occurring in 10+ plots									
complementary log-log	1	13	10	65	11	15			**115**
logistic	68	142	16	153	8	11	7	2	**407**
All	**69**	**155**	**26**	**218**	**19**	**26**	**7**	**2**	**522**
Species occurring in 20+ plots									
complementary log-log		2	9	43	8	12			**74**
logistic	61	94	14	117	6	9	7	2	**310**
All	**61**	**96**	**23**	**160**	**14**	**21**	**7**	**2**	**384**

Table 26a Countryside Survey - comparison of model types: All vegetation classes combined

	Best fitting model								
	d	**a,d**	**b,d**	**c,d**	**a,b,d**	**a,c,d**	**b,c,d**	**a,b,c,d**	**All**
Species occurring in 3+ plots									
complementary log-log	6	31	33	121	41	5	13	15	**265**
logistic	9	41	56	109	42	2	10	10	**279**
All	**15**	**72**	**89**	**230**	**83**	**7**	**23**	**25**	**544**
Species occurring in 10+ plots									
complementary log-log			5	90	40		12	8	**155**
logistic		14	21	75	42	1	9	6	**168**
All		**14**	**26**	**165**	**82**	**1**	**21**	**14**	**323**
Species occurring in 20+ plots									
complementary log-log				62	37		12	4	**115**
logistic		3	4	55	42	1	9	3	**117**
All		**3**	**4**	**117**	**79**	**1**	**21**	**7**	**232**

Table 26b Countryside Survey - comparison of model types: Vegetation classes analysed separately

	Best fitting model								
	d	**a,d**	**b,d**	**c,d**	**a,b,d**	**a,c,d**	**b,c,d**	**a,b,c,d**	**All**
Species occurring in 3+ plots									
complementary log-log	15	96	76	273	83	5	15	19	**582**
logistic	47	88	150	258	39	15	23	43	**663**
All	**62**	**184**	**226**	**531**	**122**	**20**	**38**	**62**	**1245**
Species occurring in 10+ plots									
complementary log-log		6	16	188	81		15	5	**311**
logistic	2	24	51	151	39	1	6	20	**294**
All	**2**	**30**	**67**	**339**	**120**	**1**	**21**	**25**	**605**
Species occurring in 20+ plots									
complementary log-log			4	131	67		15	4	**221**
logistic		6	11	94	39	1	5	12	**168**
All		**6**	**15**	**225**	**106**	**1**	**20**	**16**	**389**

Table 27a ADAS ESA - comparison of model types: All sites combined, all nest sizes

	Best fitting model

	Best fitting model								
	d	**a,d**	**b,d**	**c,d**	**a,b,d**	**a,c,d**	**b,c,d**	**a,b,c,d**	**All**
Species occurring in 3+ plots									
complementary log-log	9	15		55	7	22	2		**110**
logistic	19	23	4	80	4	87			**217**
All	**28**	**38**	**4**	**135**	**11**	**109**	**2**		**327**
Species occurring in 10+ plots									
complementary log-log		5		41	6	20	2		**74**
logistic	1	11	2	63	2	80			**159**
All	**1**	**16**	**2**	**104**	**8**	**100**	**2**		**233**
Species occurring in 20+ plots									
complementary log-log		1		29	1	18	2		**51**
logistic		2		50	1	73			**126**
All		**3**		**79**	**2**	**91**	**2**		**177**

Table 27b ADAS ESA - comparison of model types: Sites analysed separately

	Best fitting model								
	d	**a,d**	**b,d**	**c,d**	**a,b,d**	**a,c,d**	**b,c,d**	**a,b,c,d**	**All**
Species occurring in 3+ plots									
complementary log-log	39	95	5	265	16	38	13	1	**472**
logistic	110	114	21	365	15	93	6	5	**729**
All	**149**	**209**	**26**	**630**	**31**	**131**	**19**	**6**	**1201**
Species occurring in 10+ plots									
complementary log-log	1	9	1	183	6	27	12		**239**
logistic	18	27	5	226	9	61	4		**350**
All	**19**	**36**	**6**	**409**	**15**	**88**	**16**		**589**
Species occurring in 20+ plots									
complementary log-log		1		105		15	8		**129**
logistic	5	4	2	125	3	44	2		**185**
All	**5**	**5**	**2**	**230**	**3**	**59**	**10**		**314**

Table 28a ADAS ESA - comparison of model types: All sites combined, nests 1 m^2 and below

	Best fitting model								
	d	**a,d**	**b,d**	**c,d**	**a,b,d**	**a,c,d**	**b,c,d**	**a,b,c,d**	**All**
Species occurring in 3+ plots									
complementary log-log	3	21	1	21	5	14	2	1	**68**
logistic	6	40	4	47	3	66			**166**
All	**9**	**61**	**5**	**68**	**8**	**80**	**2**	**1**	**234**
Species occurring in 10+ plots									
complementary log-log		6		12	5	14		1	**38**
logistic		24	1	31	3	64			**123**
All		**30**	**1**	**43**	**8**	**78**		**1**	**161**
Species occurring in 20+ plots									
complementary log-log		1		8	2	14		1	**26**
logistic		11		20	3	61			**95**
All		**12**		**28**	**5**	**75**		**1**	**121**

Table 28b ADAS ESA - comparison of model types: Sites analysed separately

	Best fitting model								
	d	**a,d**	**b,d**	**c,d**	**a,b,d**	**a,c,d**	**b,c,d**	**a,b,c,d**	**All**
Species occurring in 3+ plots									
complementary log-log	12	72	5	78	19	21	3	5	**215**
logistic	40	160	24	198	10	40	2	13	**487**
All	**52**	**232**	**29**	**276**	**29**	**61**	**5**	**18**	**702**
Species occurring in 10+ plots									
complementary log-log		6	1	37	17	20		3	**84**
logistic	9	68	8	124	10	30	1	3	**253**
All	**9**	**74**	**9**	**161**	**27**	**50**	**1**	**6**	**337**
Species occurring in 20+ plots									
complementary log-log		1		14	9	16		2	**42**
logistic	5	24		62	10	27			**128**
All	**5**	**25**		**76**	**19**	**43**		**2**	**170**

Table 29a ADAS ESA - comparison of model types: All sites combined, nests above 1 m^2

	Best fitting model								
	d	**a,d**	**b,d**	**c,d**	**a,b,d**	**a,c,d**	**b,c,d**	**a,b,c,d**	**All**
Species occurring in 3+ plots									
complementary log-log	8	38	22	54	9	2	3	8	**144**
logistic	24	53	24	65		8	1	8	**183**
All	**32**	**91**	**46**	**119**	**9**	**10**	**4**	**16**	**327**
Species occurring in 10+ plots									
complementary log-log		21	12	51	9	1		5	**99**
logistic	2	45	21	58		4	1	3	**134**
All	**2**	**66**	**33**	**109**	**9**	**5**	**1**	**8**	**233**
Species occurring in 20+ plots									
complementary log-log		7	5	46	9	1		4	**72**
logistic		35	16	48		3		3	**105**
All		**42**	**21**	**94**	**9**	**4**		**7**	**177**

Table 29b ADAS ESA - comparison of model types: Sites analysed separately

	Best fitting model								
	d	**a,d**	**b,d**	**c,d**	**a,b,d**	**a,c,d**	**b,c,d**	**a,b,c,d**	**All**
Species occurring in 3+ plots									
complementary log-log	49	195	163	100		6	2	19	**534**
logistic	161	152	144	97		31	34	48	**667**
All	**210**	**347**	**307**	**197**		**37**	**36**	**67**	**1201**
Species occurring in 10+ plots									
complementary log-log	13	61	87	72		1	1	13	**248**
logistic	60	119	77	57		1	12	15	**341**
All	**73**	**180**	**164**	**129**		**2**	**13**	**28**	**589**
Species occurring in 20+ plots									
complementary log-log	8	22	34	44				5	**113**
logistic	27	85	39	36			7	7	**201**
All	**35**	**107**	**73**	**80**			**7**	**12**	**314**

Goodness of fit

4.20 The analyses presented above have examined which model is best at representing observed frequency-area curves. However, the question of goodness of fit has not yet been considered. One possible problem is introduced by the non-linear nature of the curves. This means that function optimisation algorithms must be used to obtain parameter values and such algorithms may not always converge to optimum values. In studies of individual species the behaviour of the fitting algorithm can be examined and various strategies applied to try and insure that the desired results are obtained. In the current study the vast number of models fitted makes such individual attention impossible. The consequence is an additional source of variation that may make results less clear cut than they should be. A range of individual fits have been examined in detail, however, and suggest that this problem has little overall effect.

4.21 A model may be the best fit to a particular set of data yet be little better than many of its competitors. Equally the best fitting model may not be an adequate representation of the data. The models used here are based on the presence or absence of species in sets of nested quadrats and assume multinomial variation about the fitted proportions at each nest level. This can be used to give an indication of fit, that is, whether a particular model adequately explains the variation in the data. However it should be borne in mind that an apparent lack of fit should not of itself be taken to imply that a model is incorrect. It may simply be that the variation is not multinomial but of some other form or that extra variation is due to additional factors not considered in these studies, such as the nutrient status of individual plots, a problem known as over-dispersion (see for example, Collett 1991).

4.22 Tables 30 to 33 summarise the adequacy of the different model parameterisations, using the deviance of the fitted model as the criteria for goodness of fit. The four tables represent combinations of the two types of model and the site-specific and all sites combined datasets. Results are presented separately for species found in 3+, 10+ and 20+ to allow examination to be made of the effect of limited information. Table entries are the number of species adequately modelled by the specified model. In Table 30, for example, 483 of the 544 species found in the CS dataset in three or more quadrats are adequately modelled by the complementary log-log *cd* parameterisation.

Table 30 Adequacy of individual complementary log-log curves: All sites combined

Curve parameters	Natural England	Countryside Survey	ADAS ESA monitoring		
			all nests	small nests	large nests
Species found in 3 or more quadrats					
d	16	79	86	37	87
a, d	32	188	118	87	173
b, d	21	265	95	57	212
c, d	56	483	231	158	301
a, b, d	38	469	130	118	295
a, c, d	87	449	258	201	292
b, c, d	52	457	223	146	280
a, b, c, d	79	483	250	192	289
Num. species	127	544	327	234	327
Species found in 10 or more quadrats					
d	4	8	12	4	10
a, d	17	32	32	26	82
b, d	6	98	16	11	124
c, d	40	273	137	89	208
a, b, d	22	260	42	53	203
a, c, d	71	248	164	130	202
b, c, d	37	269	129	80	193
a, b, c, d	64	281	157	121	200
Num. species	111	323	233	161	233
Species found in 20 or more quadrats					
d	1	1	1	1	
a, d	7	6	7	8	43
b, d	2	33	1	2	83
c, d	29	190	84	57	156
a, b, d	12	178	7	25	152
a, c, d	60	171	108	92	152
b, c, d	26	190	77	49	146
a, b, c, d	53	202	102	84	150
Num. species	100	232	177	121	177

Table 31 Adequacy of individual logistic curves: Adequacy of individual logistic curves

Curve parameters	Natural England	Countryside Survey	ADAS ESA monitoring		
			all nests	small nests	large nests
Species found in 3 or more quadrats					
d	16	79	86	37	88
a, d	55	256	140	141	240
b, d	21	279	96	58	224
c, d	61	469	240	158	295
a, b, d	71	500	156	165	300
a, c, d	107	450	262	212	292
b, c, d	54	457	229	128	279
a, b, c, d	97	472	254	205	291
Num. species	127	544	327	234	327
Species found in 10 or more quadrats					
d	4	8	12	4	11
a, d	40	71	48	72	149
b, d	6	100	17	11	131
c, d	45	271	146	90	207
a, b, d	55	290	64	94	208
a, c, d	91	246	168	140	202
b, c, d	39	269	135	81	192
a, b, c, d	82	277	161	134	199
Num. species	111	323	233	161	233
Species found in 20 or more quadrats					
d	1	1	1	1	1
a, d	29	21	15	38	102
b, d	2	34	1	2	89
c, d	34	189	93	58	156
a, b, d	44	208	21	58	157
a, c, d	80	169	112	102	152
b, c, d	28	190	83	51	145
a, b, c, d	71	201	106	97	149
Num. species	100	232	177	121	177

Table 32 Adequacy of individual complementary log-log curves: Sites modelled separately

Curve parameters	Natural England	Countryside Survey	ADAS ESA monitoring		
			all nests	small nests	large nests
Species found in 3 or more quadrats					
d	162	237	751	179	530
a, d	291	544	782	365	938
b, d	230	717	764	268	1076
c, d	539	1168	1101	601	1159
a, b, d	423	1150	813	489	1154
a, c, d	619	1081	1115	645	1143
b, c, d	486	1070	1083	571	1130
a, b, c, d	557	1123	1098	624	1150
Num. species	694	1245	1201	702	1201
Species found in 10 or more quadrats					
d	47	10	139	7	83
a, d	136	66	170	58	348
b, d	103	219	152	42	498
c, d	374	549	489	247	554
a, b, d	262	529	201	152	554
a, c, d	451	496	503	284	539
b, c, d	325	524	471	232	550
a, b, c, d	397	527	486	273	544
Num. species	522	605	589	337	589
Species found in 20 or more quadrats					
d	15	1	3	1	20
a, d	59	10	9	14	147
b, d	59	73	6	3	250
c, d	262	345	228	111	290
a, b, d	160	330	22	51	292
a, c, d	325	309	236	139	281
b, c, d	222	334	212	103	290
a, b, c, d	280	346	226	132	286
Num. species	384	389	314	170	314

Table 33 Adequacy of individual logistic curves: Sites modelled separately

Curve parameters	Natural England	Countryside Survey	ADAS ESA monitoring		
			all nests	small nests	large nests
Species found in 3 or more quadrats					
d	357	239	514	212	685
a, d	578	762	771	556	1114
b, d	407	767	596	318	1130
c, d	603	1131	1106	622	1130
a, b, d	613	1183	874	623	1149
a, c, d	629	1074	1115	656	1140
b, c, d	571	1077	1099	591	1126
a, b, c, d	618	1113	1104	641	1152
Num. species	694	1245	1201	702	1201
Species found in 10 or more quadrats					
d	238	10	98	32	217
a, d	414	184	204	207	515
b, d	276	229	135	79	532
c, d	438	542	496	266	550
a, b, d	447	564	292	265	550
a, c, d	462	487	504	295	537
b, c, d	409	527	491	249	555
a, b, c, d	453	530	494	285	546
Num. species	522	605	589	337	589
Species found in 20 or more quadrats					
d	195	1	22	6	96
a, d	297	51	48	84	263
b, d	219	83	31	17	278
c, d	324	338	240	127	287
a, b, d	322	362	89	123	291
a, c, d	335	302	242	143	279
b, c, d	303	337	235	113	292
a, b, c, d	331	347	236	139	286
Num. species	384	389	314	170	314

4.23 The results are relatively similar for all of the tables. Overall a relatively high proportion of species are adequately represented by these models suggesting that the models are suitable for the majority of species and that the main influence on species occurrence is quadrat size. The logistic model is marginally better than the complementary log-log, particularly for the NE and full ADAS datasets.

4.24 As expected models with higher numbers of parameters fit the data better but there are variations within this pattern. For the NE data, the complete ADAS data and the ADAS data from the smaller nests, the *cd* and *acd* parameterisations stand out as being more adequate representations of the data. For the CS data and the ADAS data from large nests the three and four parameter models are roughly equal in their ability to explain the data. Of the remaining parameterisations the *cd* parameterisation is much better than the other two and one parameter models and as adequate as the three parameter models. The *cd* parameterisation is judged adequate for a higher proportion of species in the site-specific analyses. This could be a reflection of the greater sensitivity of the model fit statistic when based on larger amounts of data but, since it is not found for all parameterisations, is more likely to be a reflection of the need for a more complex model with heterogeneous data.

4.25 A final factor affecting model fit is, as described above, the amount and extent of the data. Models fit best when the species involved is found in a reasonably large number of quadrats spread over a wide range of nest sizes. Discrepancies from this pattern can reduce the ability to discriminate between models and will affect the accuracy with which chosen models can be used for prediction or extrapolation.

Examples

4.26 In this section we illustrate model fit with examples from each of the datasets. Both typical results and examples of the effect of data extremes are presented. These shed considerable light on the results tabulated above and clarify the differences found between the three datasets. Not all curves are apparent in each plot, though all have been plotted. Where the parameters of one curve are a subset of the parameters of another, and the extra parameters do not improve the fit, the curve plotted second will overlay that plotted first.

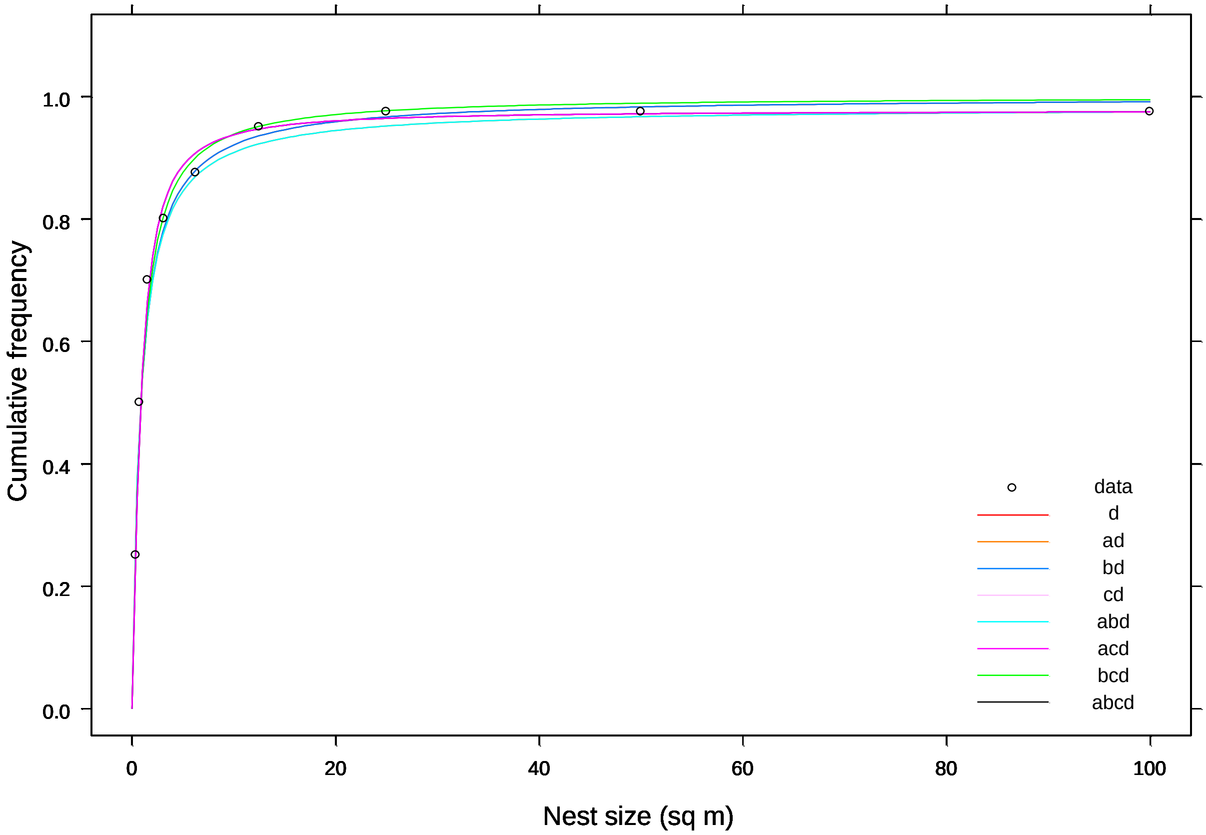


Figure 6 Curve fitting for Natural England data: logistic model - Wylye and Church Dean Down *Succisa pratensis*

4.27 Figure 6 shows a species, *Succisa pratensis*, occurring at high density at one of the NE sites. The observed frequency rises steeply at small nest sizes before flattening out at a frequency close to 1.0. This is a typical pattern for high density species for which most observations occur in the first few, nest sizes. The fitting procedure reacts by choosing very large *c* and *d* values. All curves fit well in this situation and any could be used to represent the data.

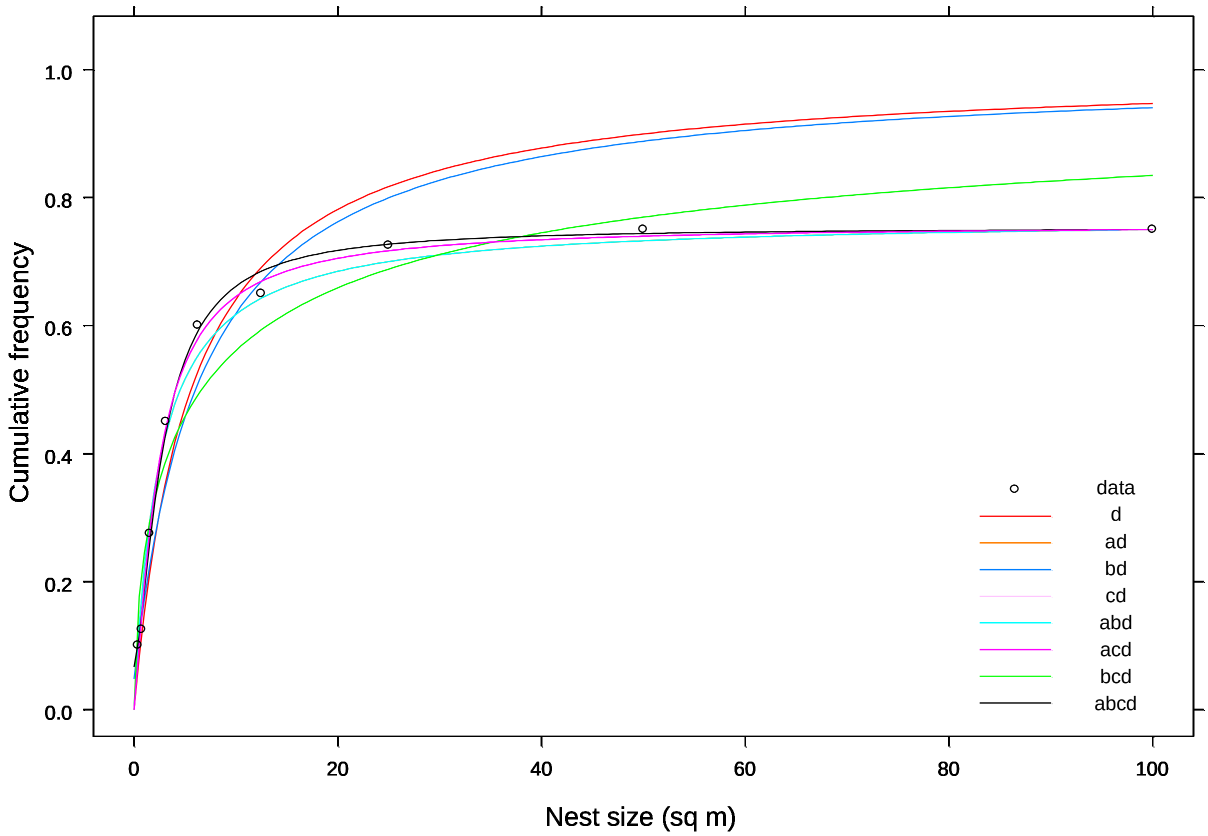


Figure 7 Curve fitting for Natural England data: logistic model - Knighton Bank and Wood *Koeleria macrantha*

4.28 Figure 7 shows a different scenario. The *Koeleria macrantha* curve also rises sharply but this time flattens out at a frequency below 0.8. Only those parameterisations containing the a parameter are capable of accurately representing this data and most do this reasonably well, though the *acd* parameterisation performs best. The *d* and *bd* parameterisations, because they are forced to rise to a maximum frequency of 1.0, underestimate the frequency in the smaller nests and overestimate the frequencies in the higher nests. The *cd* and *bcd* parameterisations (superimposed), which also have to rise to a maximum value of 1.0, make use of the c parameter to counteract this problem and hence produce a better overall fit. They still, however, underestimate the sharpness of the shoulder in the data curve and somewhat overestimate the frequency at the larger nest sizes.

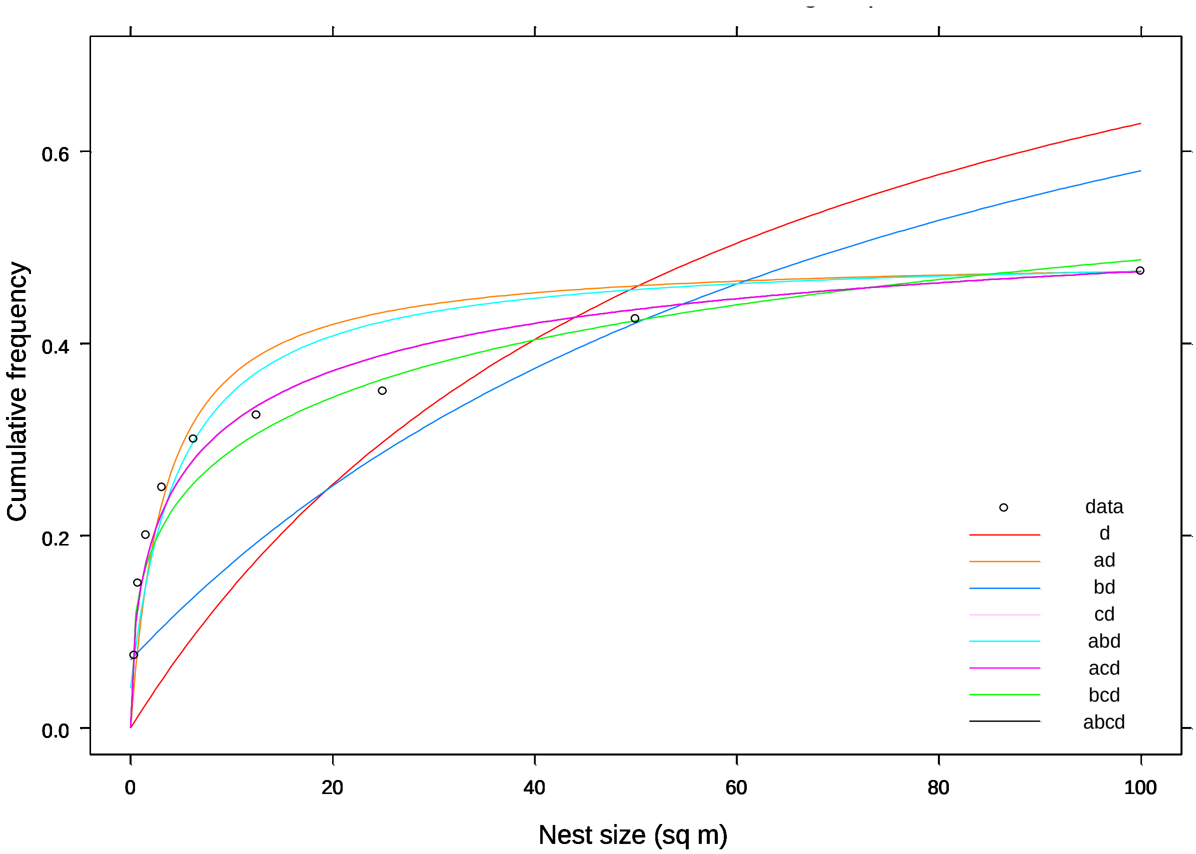


Figure 8 Curve fitting for Natural England data: logistic model - Knighton Down and Wood *Medicago lupulina*

4.29 Figure 8 shows a more extreme example. In this case the *ad* and *abd* parameterisations are unable to adapt to the shape of the observed data despite including the *a* parameter and both the *a* and *c* parameters are needed to give an adequate fit.

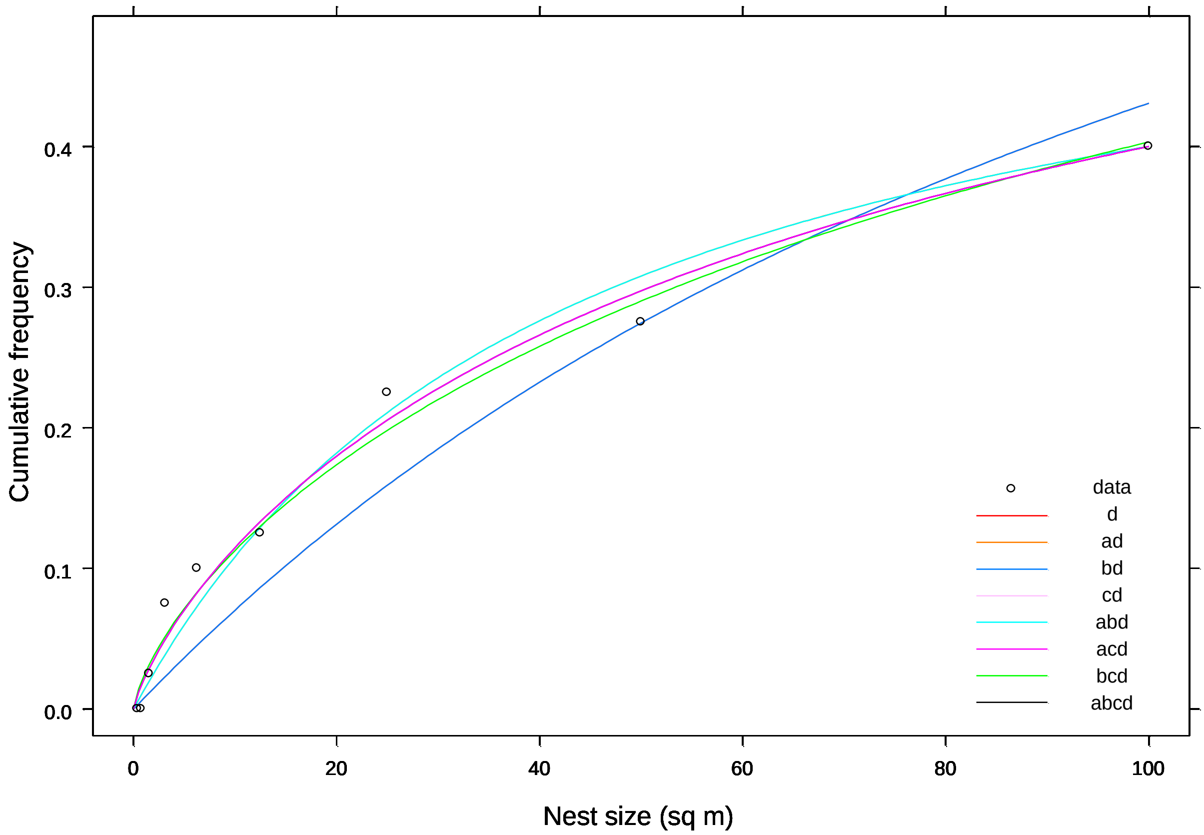


Figure 9 Curve fitting for Natural England data: logistic model - Wylye and Church Dean Down *Filipendula vulgaris*

4.30 Figure 9 shows an example where frequency increases steadily with nest size. By the largest nest size the data has not yet formed a shoulder as in the previous figures and shows no sign of flattening out. With the exception of the *d* and *bd* parameterisations (superimposed) all curves fit reasonably well. However, they are unlikely to agree well outside the range of the data so their use for prediction to larger nest sizes could be problematical. The difficulty is that the variability of the observed values about the fitted curves is sufficiently great for the behaviour of the species at nest sizes larger than those used in the data to be ill-defined.

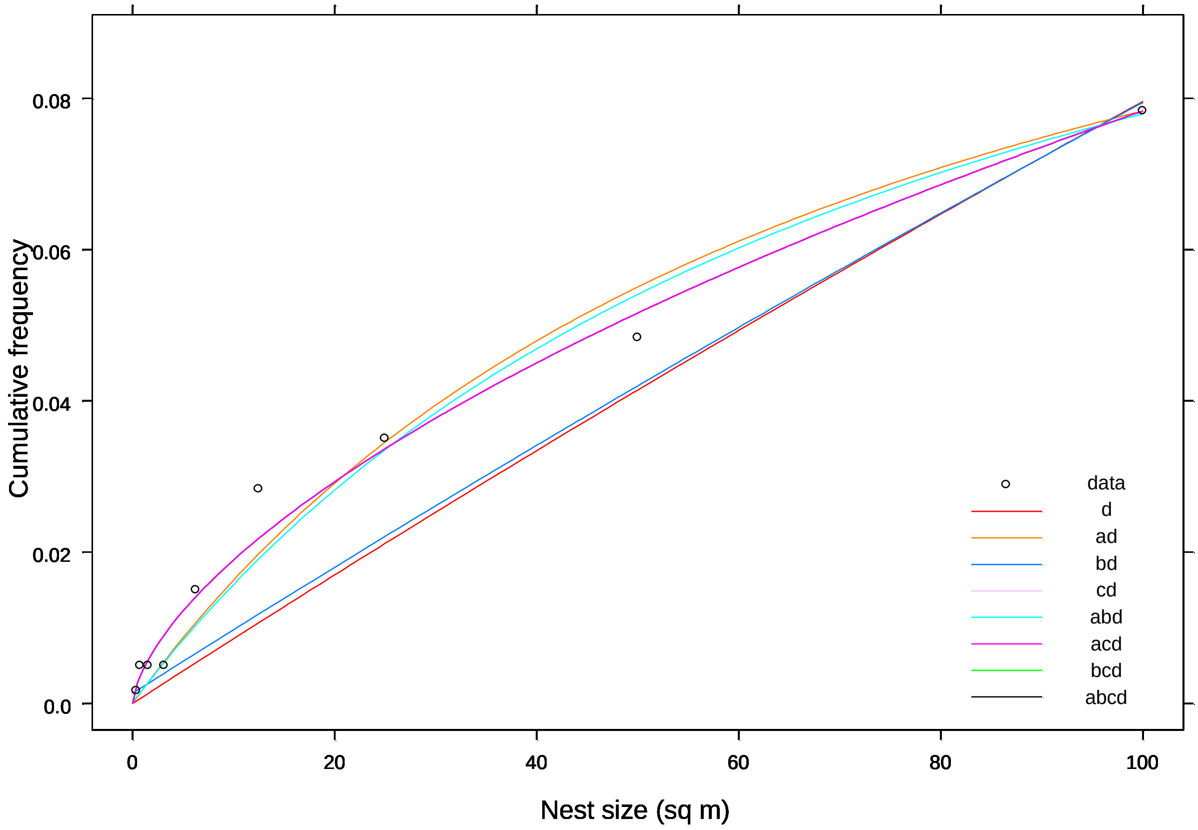


Figure 10 Curve fitting for Natural England data: logistic model all sites combined *Heracleum sphondylium*

4.31 Figure 10 shows a similar but even more extreme example. Here the maximum frequency reached by the largest nest size is only 0.08, leaving enormous scope for behaviour outside the range of the data.

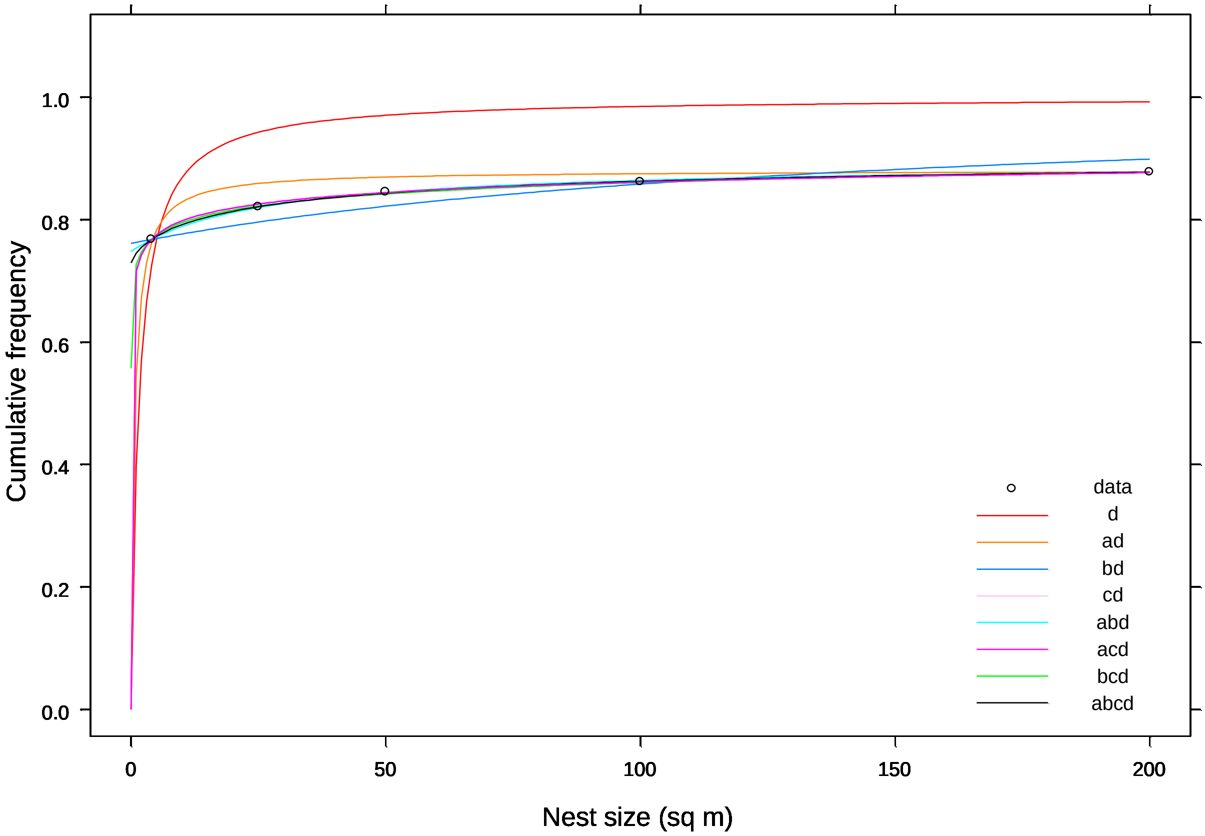


Figure 11 Curve fitting for Countryside Survey data: logistic model Aggregate Class 4 *Holcus lanatus*

4.32 Figure 11 shows a species with high frequency of occurrence in the CS data, Aggregate Class 4. This curve illustrates one of the reasons for the different model choices made for this dataset. Many of the species examined show this pattern, with little change in the frequency from the smallest nest size to the largest. The density of the species illustrated is such that the initial steep rise and shoulder of the frequency-area curve, such as shown in Figure 6, has occurred at nest sizes smaller than the smallest used by CS (4 m^2). Furthermore the small number of nests used means that there are few points at the bottom of the area range with which to determine the shape of the curve. The result is that the fitting algorithm can achieve a reasonable fit to the observed data using any of the parameters *a, b* or *c* in addition to *d*. Only the single parameter curve can not achieve a good fit.

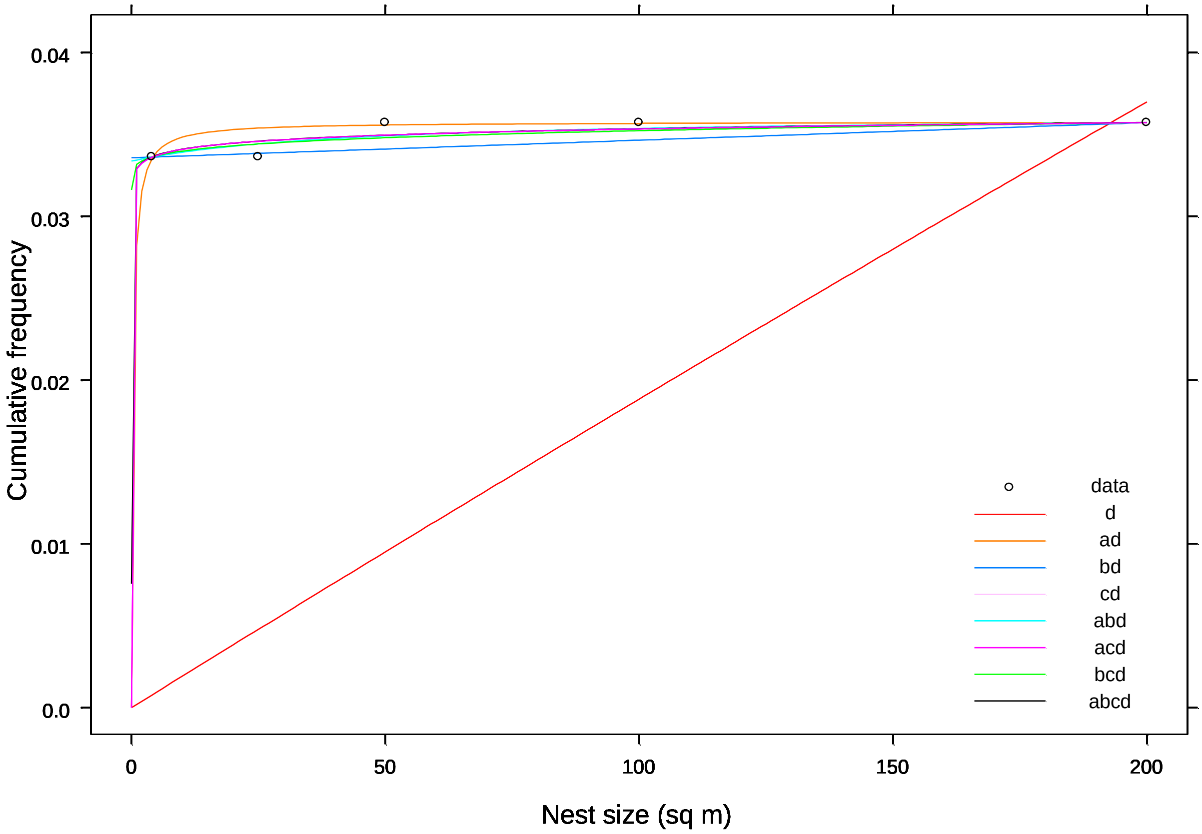


Figure 12 Curve fitting for Countryside Survey data: logistic model Aggregate Class 1 *Geranium molle*

4.33 Figure 12 shows a similar example but at very much lower density.

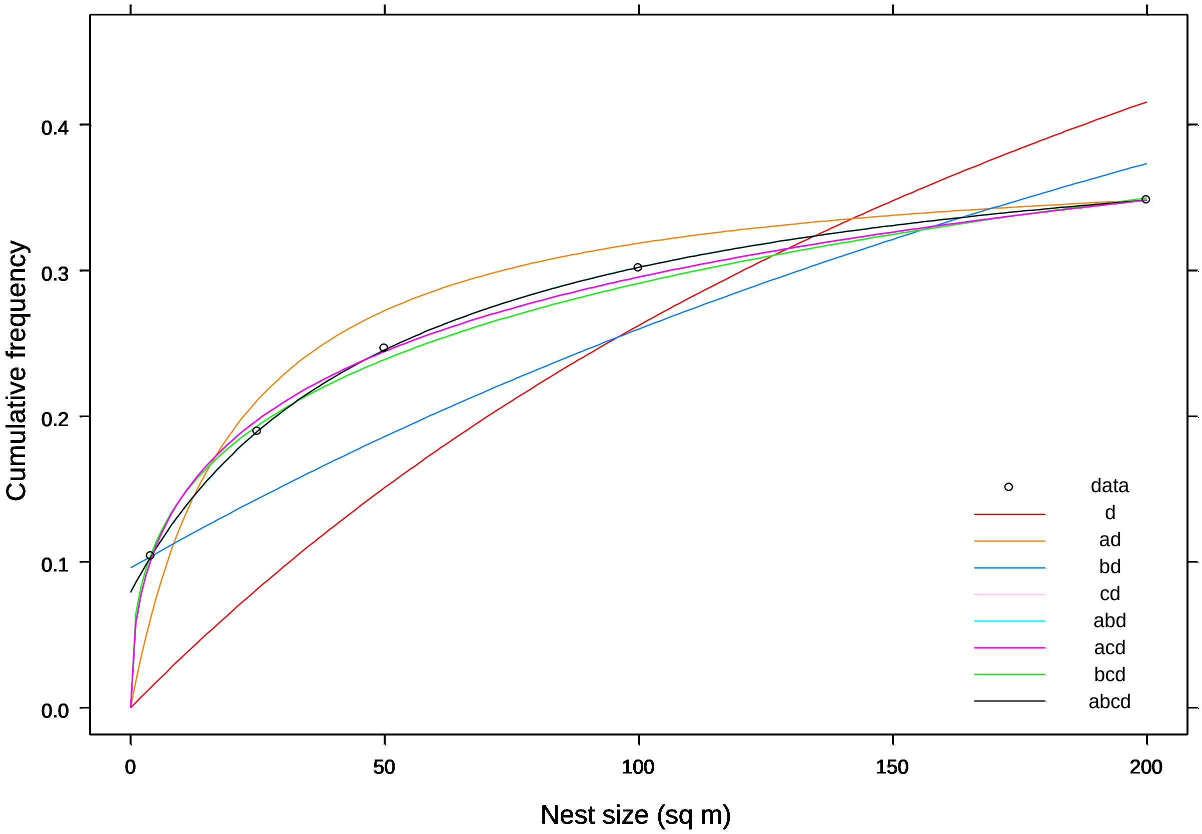


Figure 13 Curve fitting for Countryside Survey data: logistic model Aggregate Class 4 *Cirsium vulgare*

4.34 Figure 13 shows an example in which the rise in frequency of the species, *Cirsium vulgare*, does occur within the range of nest sizes used. The result is that a good fit to the data is achieved by a range of models.

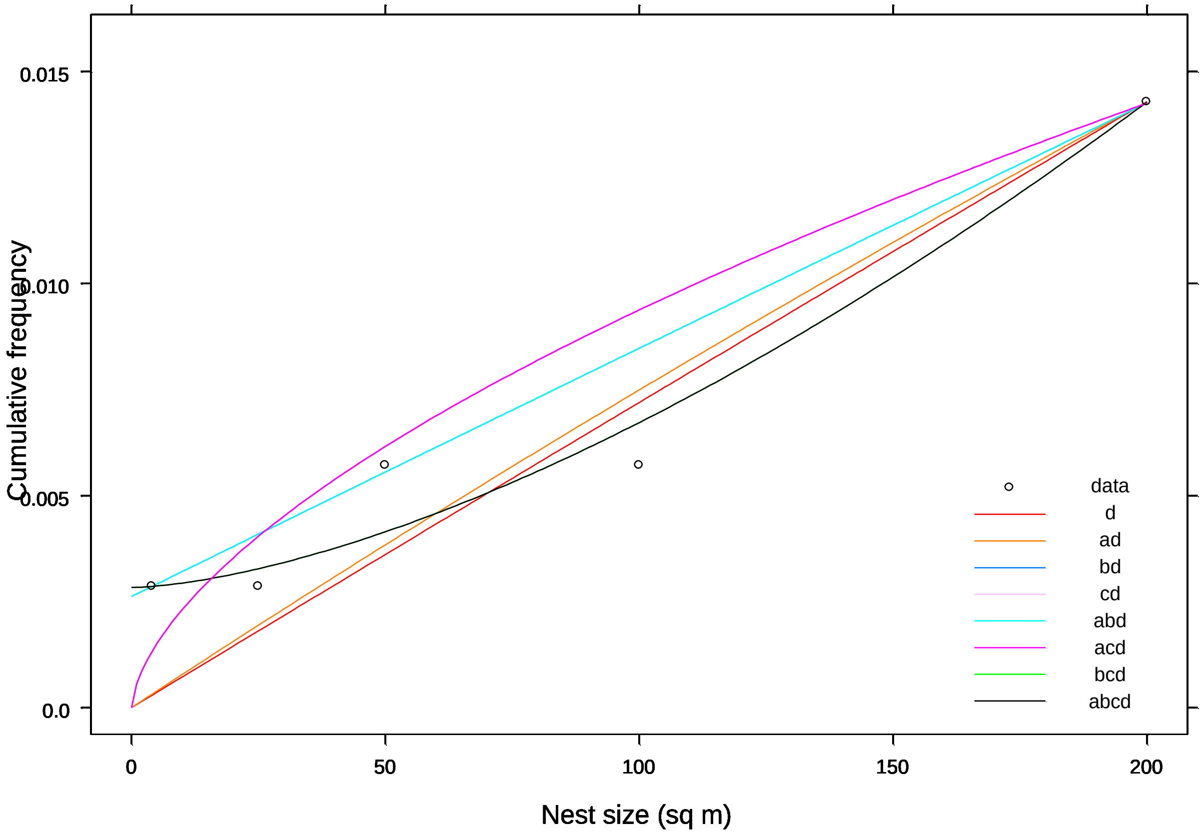


Figure 14 Curve fitting for Countryside Survey data: logistic model Aggregate Class 7 *Melampyrum pratense*

4.35 Accurate curve fitting is particularly difficult for low density species with this dataset. The interaction between the extra variability caused by the low number of observations of such species and the small number of nest sizes acts to produce questionable results. In Figure 14 none of the curves fit very well and extrapolation outside the range of the data would be very unreliable.

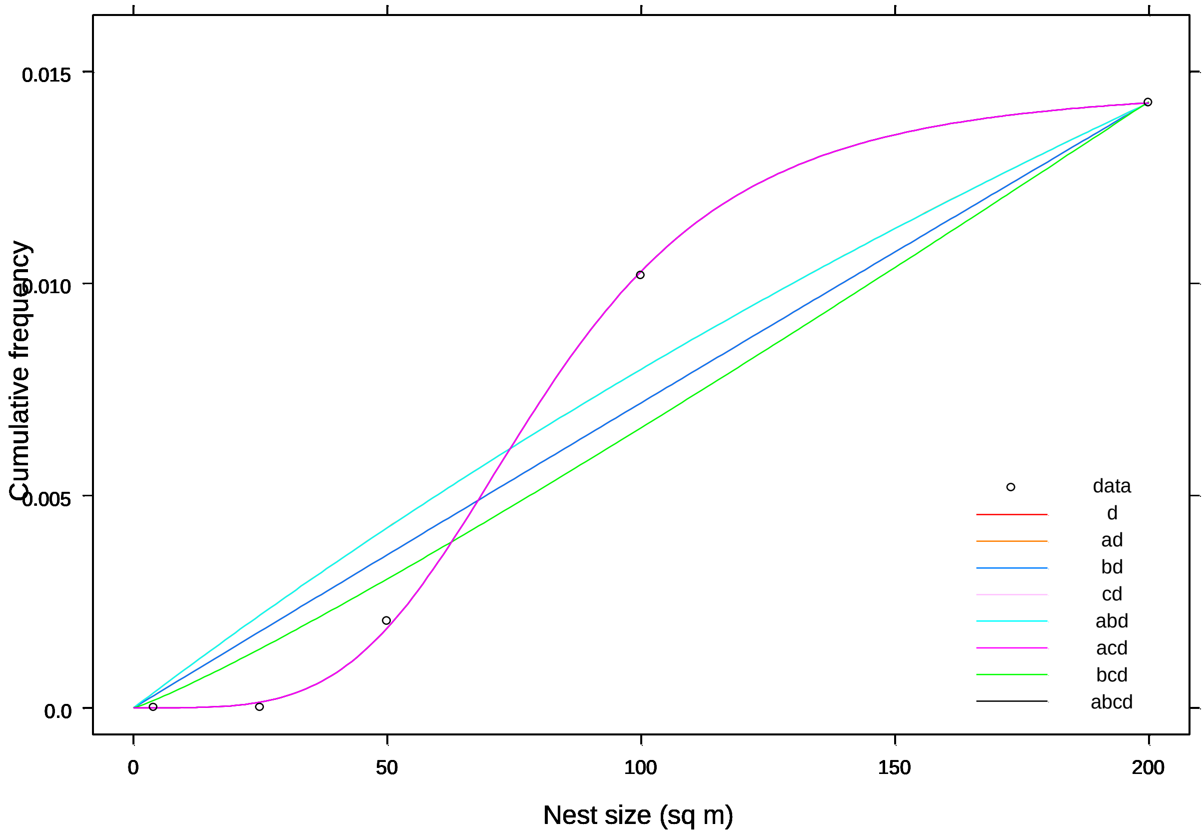


Figure 15 Curve fitting for Countryside Survey data: logistic model Aggregate Class 4 *Geranium robertianum*

4.36 In Figure 15 the *acd* and *abcd* parameterisations (superimposed) appear to fit the observed data remarkably well but this result is almost certainly due to the small number of data points compared to the number of parameters.

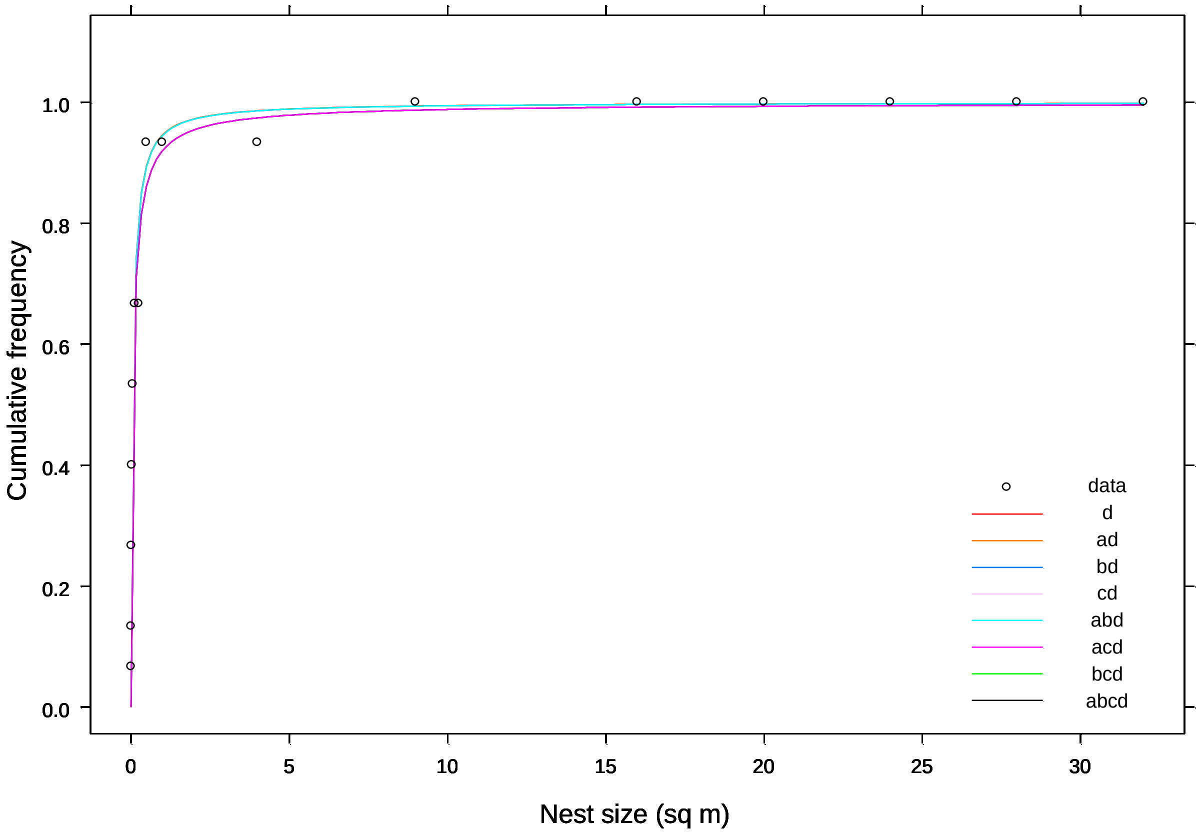


Figure 16 Curve fitting for ADAS data: logistic model Clun ESA B8 *Cerastium fontanum*

4.37 The wide size range and large number of nest sizes used in this dataset means that in general the fitted curves are more reliable and have much wider applicability. Figure 16 shows fitted curves for a high-density species. All curves produce a good fit to the data. At intermediate densities, as in Figures 17 and 18 more discrimination between curves is possible and it becomes clear that the *cd* and *acd* parameterisations fit much better. Even at very low densities (Figures 19 and 20) there is good agreement between the different parameterisations and, unlike the NE data, the range of nest sizes is sufficiently great that comparison with other datasets is likely to involve interpolation rather than extrapolation, a much more accurate process.

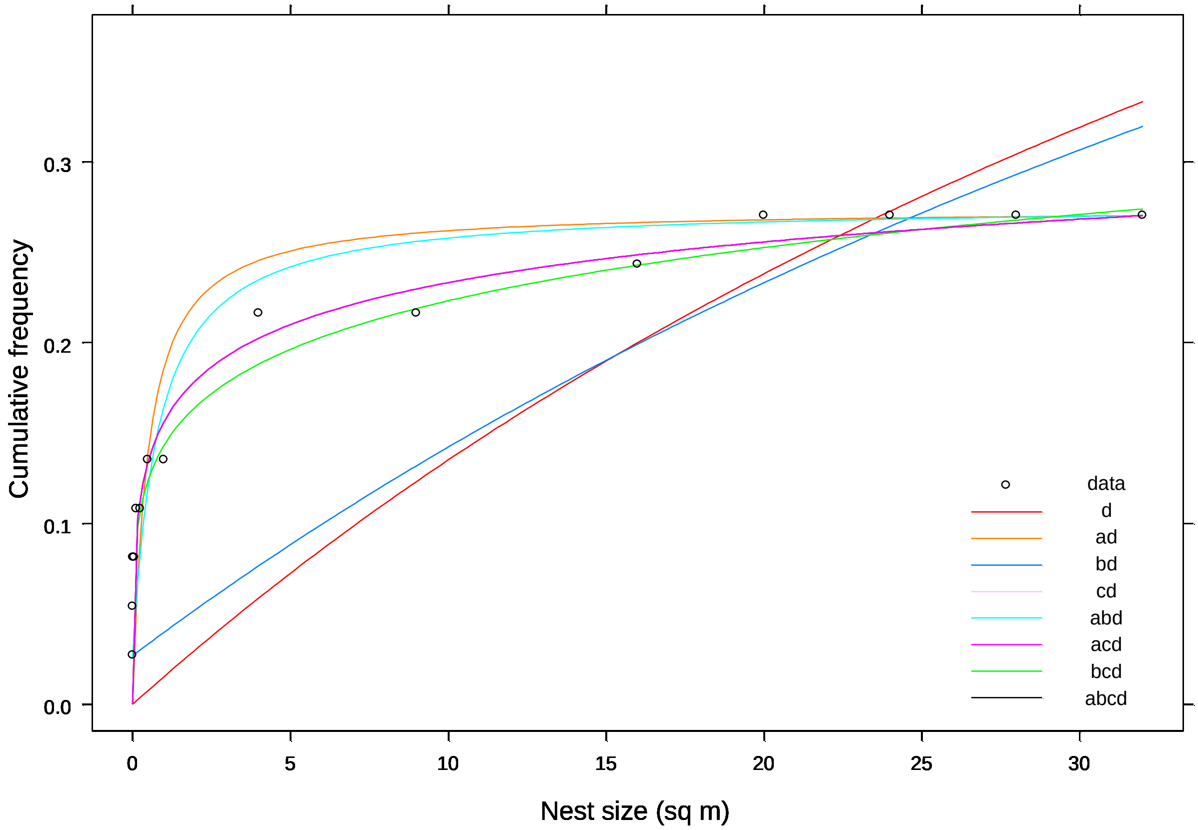


Figure 17 Curve fitting for ADAS data: logistic model Blackdown Hills B1 *Agrostis vinealis*

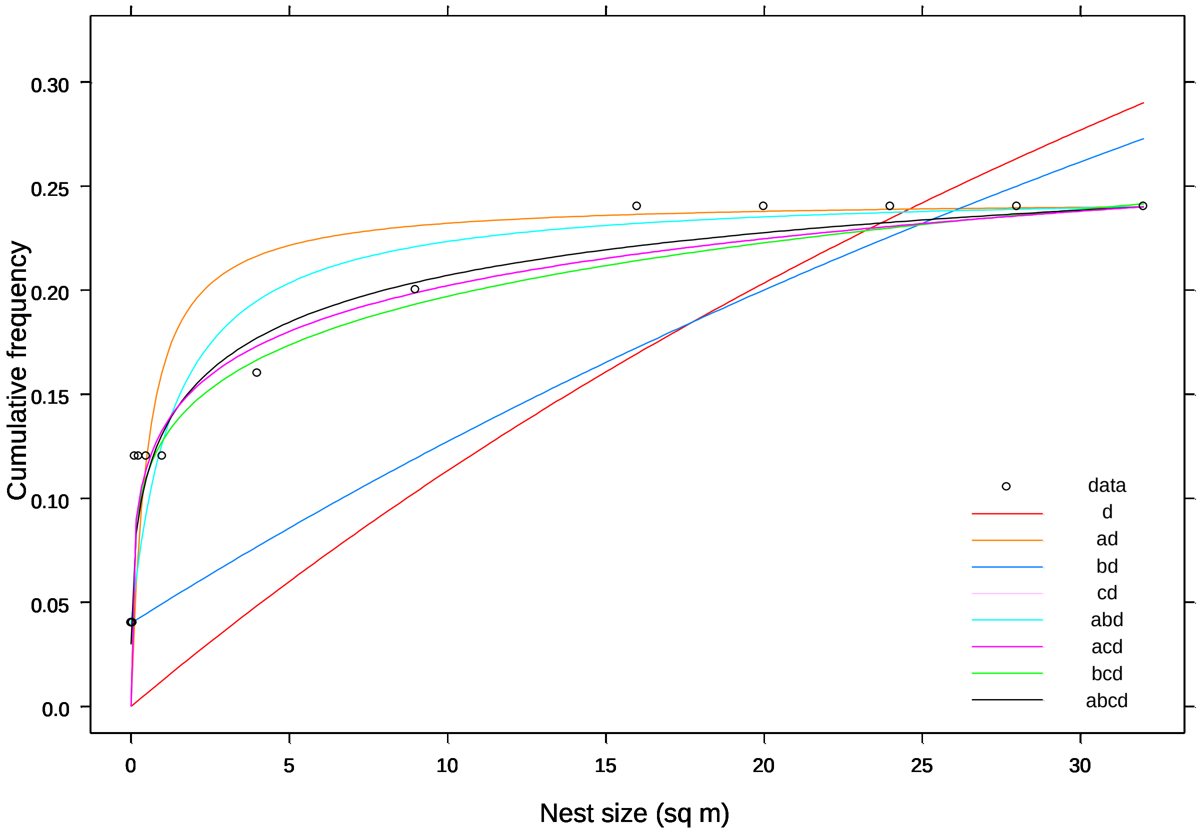


Figure 18 Curve fitting for ADAS data: logistic model Exmoor B1 *Ajuga reptans*

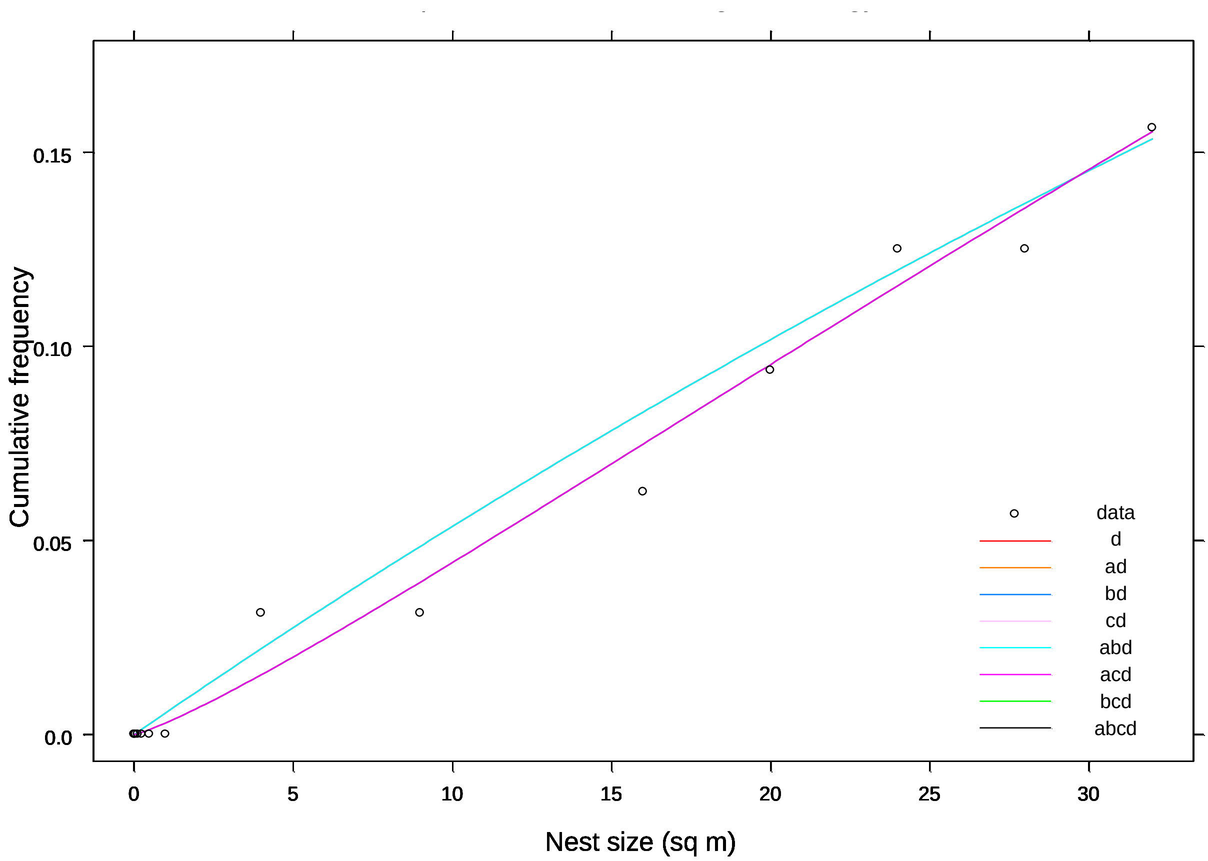


Figure 19 Curve fitting for ADAS data: logistic model Shropshire Hills B1 *Crataegus monogyna*

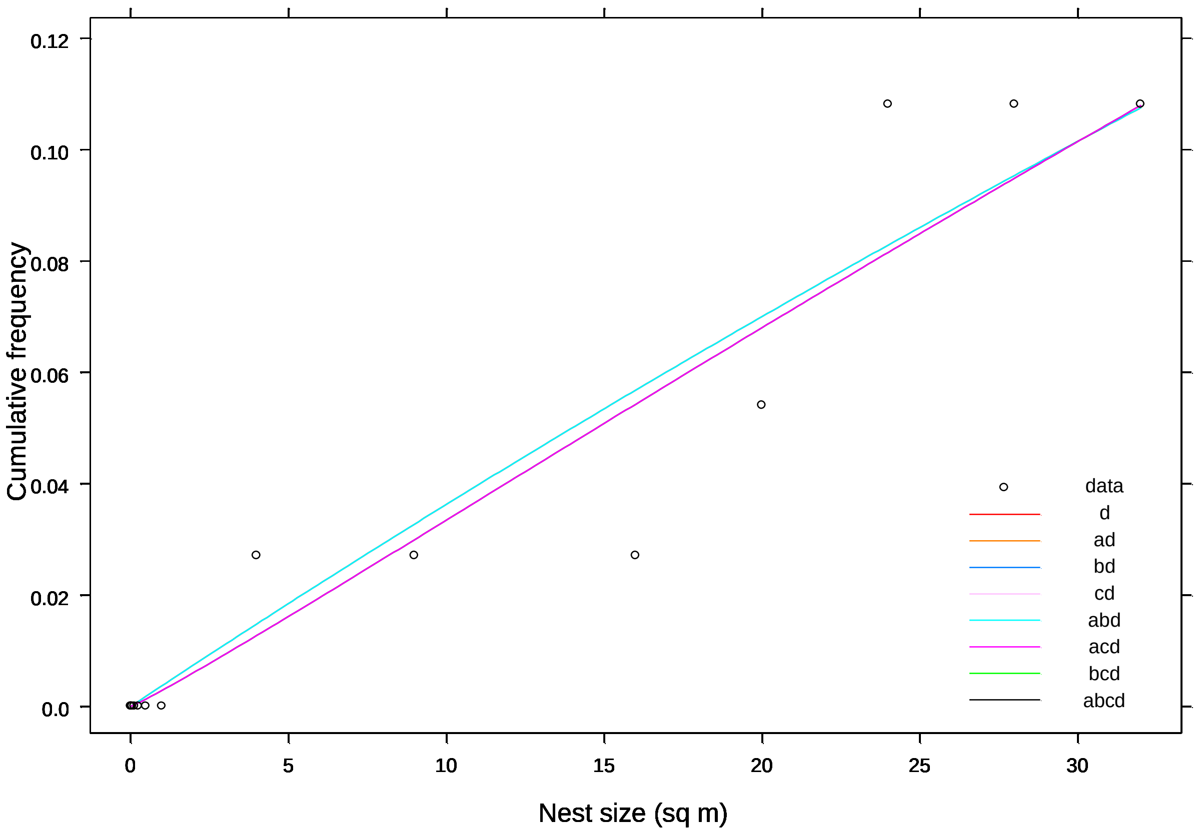


Figure 20 Curve fitting for ADAS data: logistic model Blackdown Hills B1 *Epilobium palustre*

Overview

4.38 The analyses reported in this section lead to a number of interesting conclusions. Firstly the range of models examined here provides an adequate representation of the majority of the species studied for all of the datasets. Moreover there is clear evidence that just two of the parameterisations studied are capable of representing the frequency-area relationships of the majority of species. Overall the most suitable model is the logistic *cd* parameterisation but the logistic *acd* is necessary for adequate representation of some species, particularly when the number of nests is high and their range extends close to zero size.

4.39 Secondly the same models were appropriate whether the data represented homogeneous vegetation, such as the individual NE sites, or very heterogeneous vegetation such as that from the complete CS dataset. Apparent differences between the models fitted to the different datasets arose from the different range of areas covered and the different number of nests used.

4.40 Thirdly, to obtain accurate identification and representation of a frequency-area curve requires a reasonably large number of quadrats spread over a wide range of nest sizes. Ideally species to be modelled should be observed in at least twenty plots and over six or more different nest sizes. The range of nest sizes is also important. Many species appear to exhibit a steep increase in frequency with area before reaching a relatively constant value. If possible the range of nest sizes should cover both parts of the curve.

4.41 In reaching these conclusions more weight has been given to the findings from the ADAS ESA dataset because of its larger number of nests, its wider range of quadrat areas and the fact that splitting the dataset into parts comparable to the other two datasets reproduced the patterns found in those datasets, suggesting that the differences were induced by insufficient data.

5 Analysis of fitted curve parameters

5.1 In this section variation in and relationships between the parameters of the fitted curves are examined. Only the logistic *acd* parameterisation is considered since this was found, in the previous section, to be the best fitting model overall. All the results described in this section refer to the site (NE and ADAS datasets) or aggregate vegetation class (CS) data.

5.2 Figure 21 shows the distribution of parameter values for the Natural England data plotted against the proportion of quadrats in which each species is found. This proportion acts as a proxy for the relative abundance of each species so that the plots should pick out any differences in the fitted curves between rare and common species.

5.3 The most striking aspect of Figure 21 is the restriction of the a parameter values to be equal to or greater than the proportion of quadrats in which species are found. This is a reflection of the nested quadrat method of data collection which only records additional species at each nest size. The frequency data for each species must increase with nest size with the result that fitted values of the a parameter are never less than the maximum frequency observed. The majority of a values fall either relatively close to the line of equality or have the value 1. This latter situation occurs when the data support the *cd* parameterisation rather than the *acd*. Thus fitted values of the a parameter tend to default to the two extremes of the possible range of values.

5.4 Values of the *c* parameter show little variation with species abundance, though there is the suggestion of an increase in values for the most common species and more variability for the less common species, most likely a reflection of the lack of sufficient data to produce an accurate fit. The median value is 0.82, less than the value of 1 that would indicate the suitability of a more parsimonious model.

5.5 Values of the *d* parameter show a steady decline with abundance indicating, not unexpectedly, that the frequency-area curves of abundant species have a greater rise of frequency with area than do less abundant species.

5.6 Figure 22 shows the relationships between the parameters. There is little relationship between the a parameter and the other two, possibly again reflecting the lack of ability of the nested quadrat methodology to provide detailed information about this parameter. There is, however a clear non-linear relationship between the c and d parameters. The area occupied by points is almost triangular except that those species with the steepest and shallowest frequency-area curves (high and low *d* values) have higher values of the c parameter suggesting that the spatial patterning of these species may be different from the majority.

5.7 Figure 23 shows the distribution of parameter values for the ADAS data. The results are very similar to those from the NE data. Again the a parameter values are either 1 or close to the maximum frequency observed and there is a steady, approximately linear, decline in the *d* parameter with abundance. The *c* parameter for this dataset, median value 0.52, again has more variability for less common species and exhibits evidence of a non-linear trend, with higher values for rare and common species.

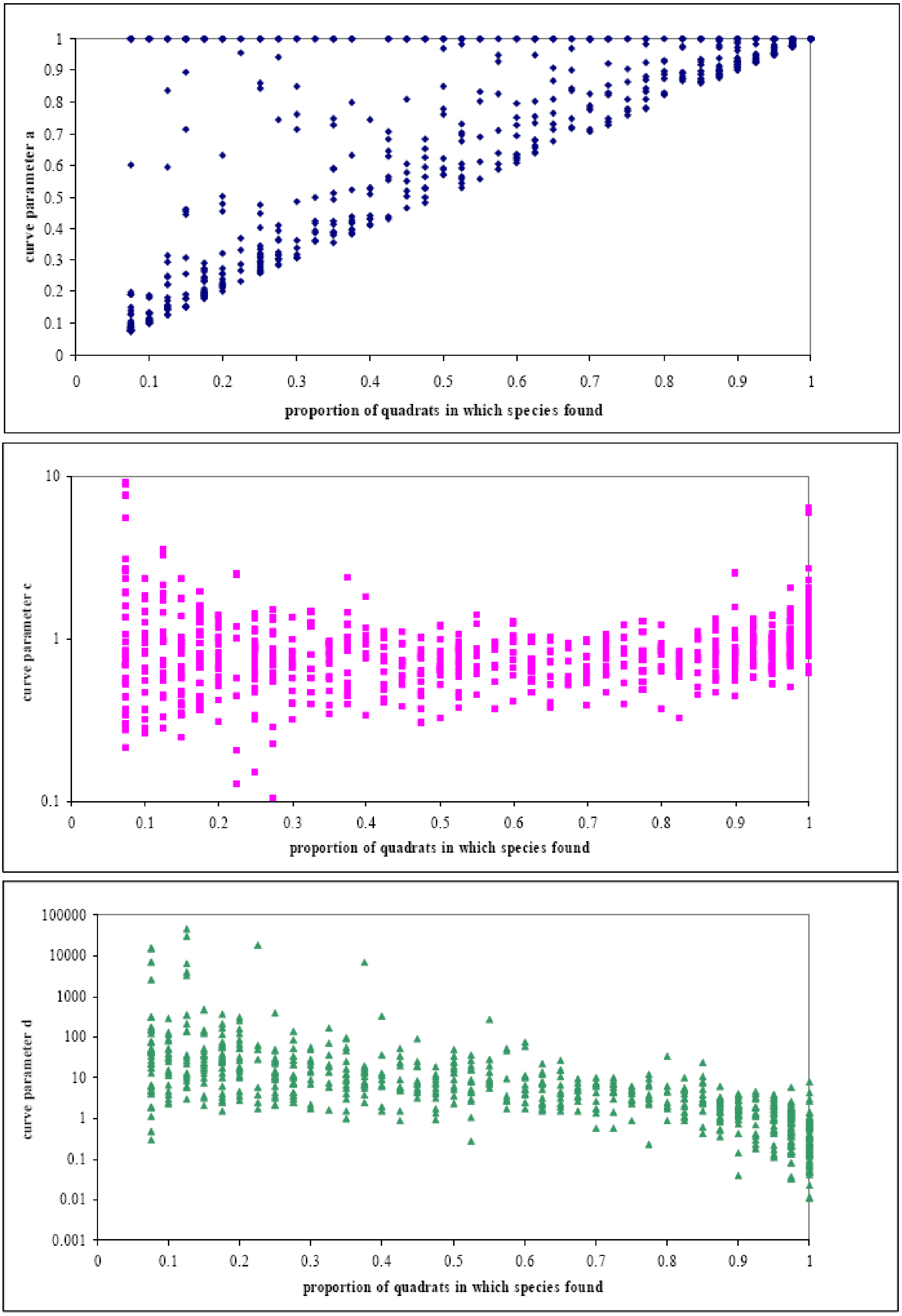


Figure 21 Natural England data – Parameters values and species frequency

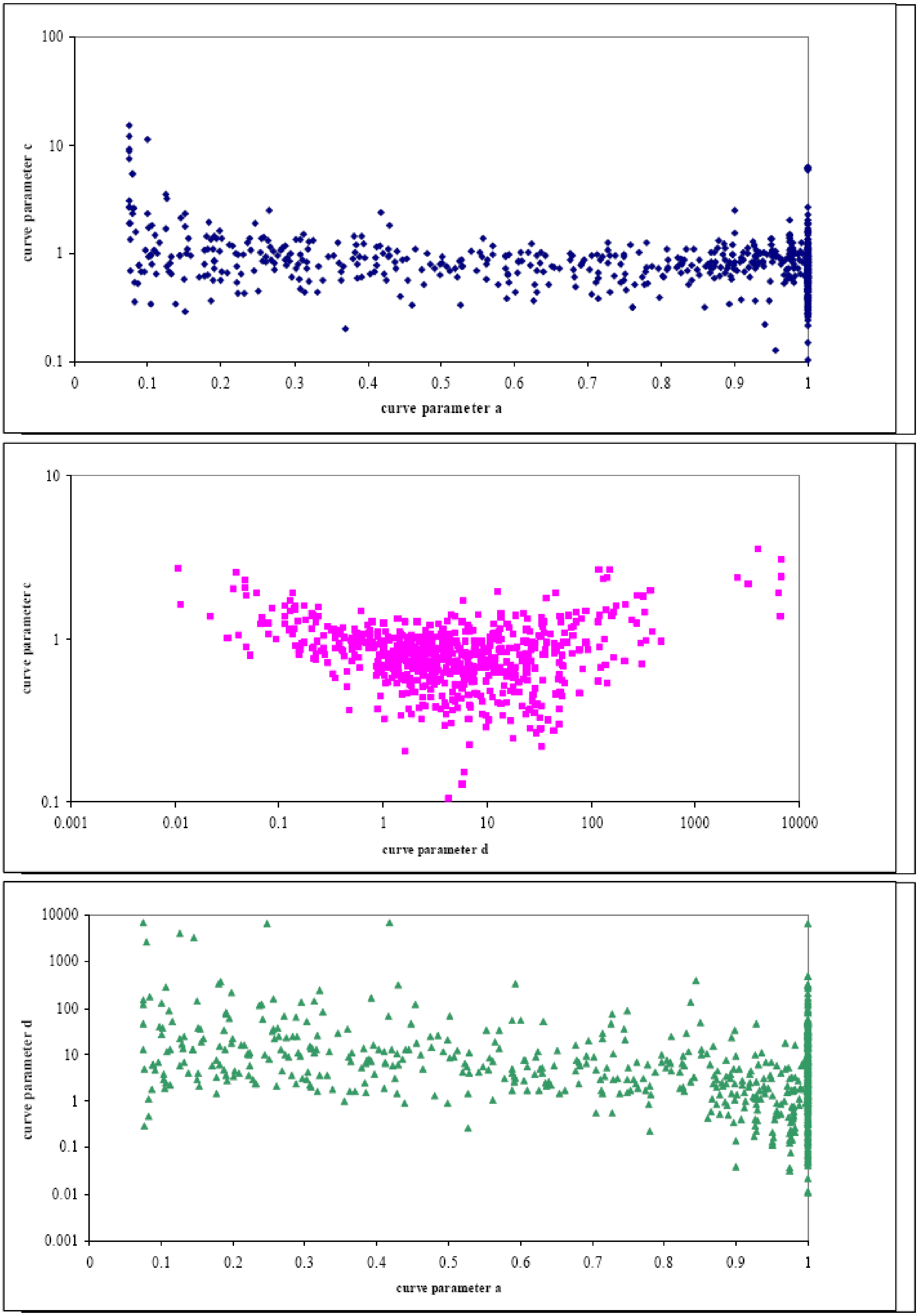


Figure 22 Natural England data – Relationship between parameters values

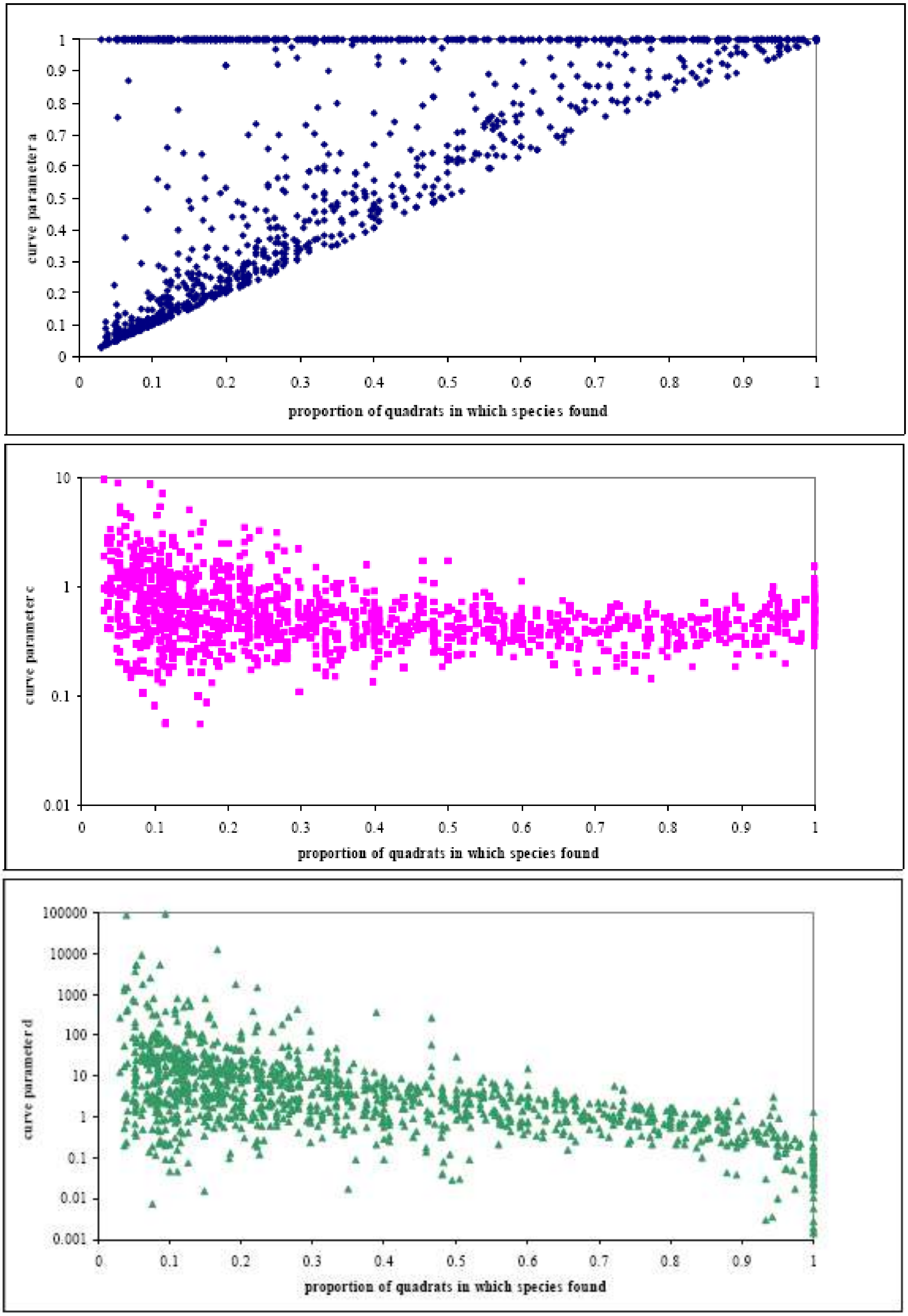


Figure 23 ADAS data – Parameters values and species frequency

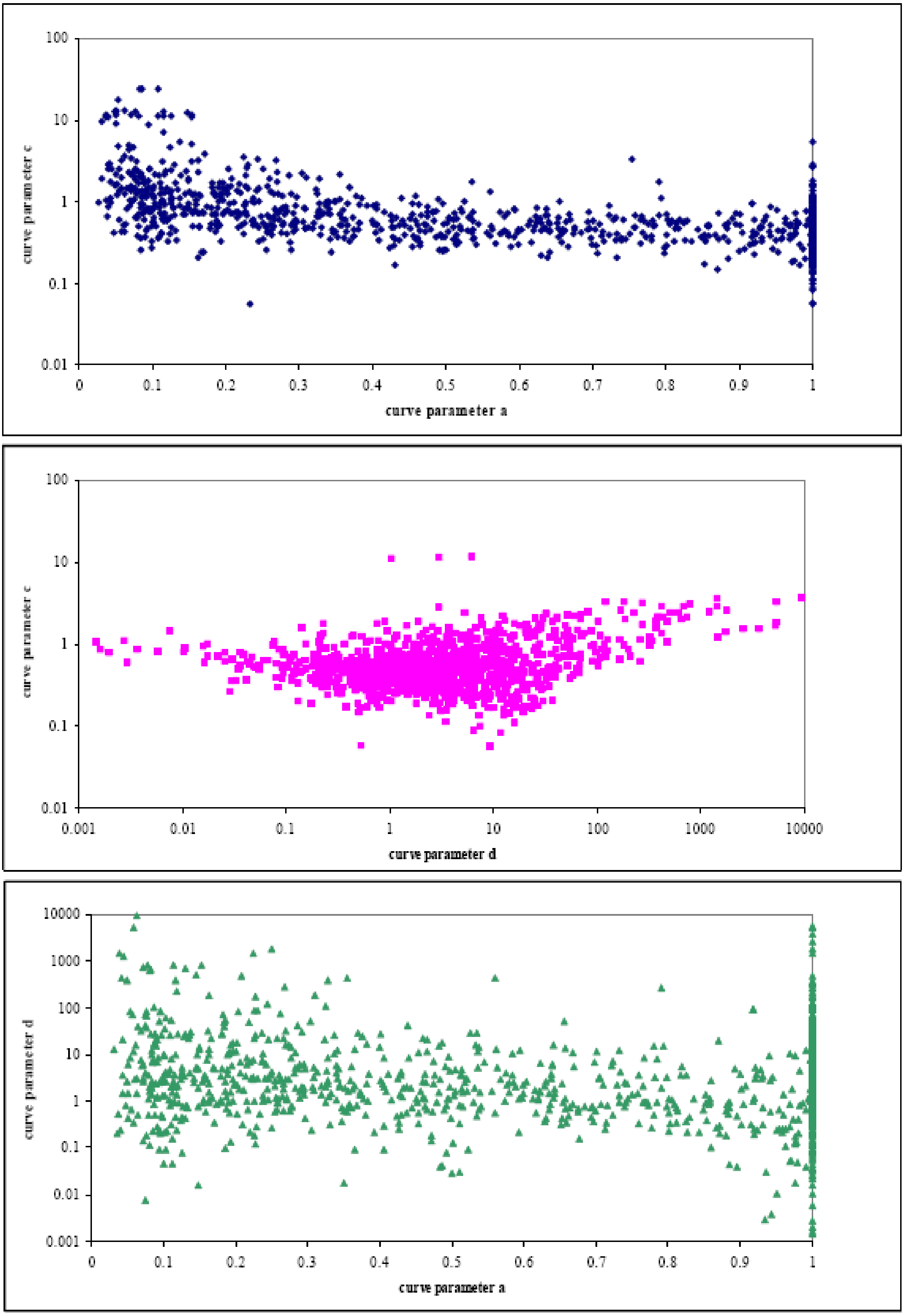


Figure 24 ADAS data – Relationship between parameters values

5.8 Figure 24 shows the relationships between the parameters for the ADAS data. There is a slight downward trend of both the *c* and *d* parameters with the a parameter but the main relationship is again the triangular relationship between the *c* and *d* parameters with an increase in *c* values at the highest *d* values.

5.9 The CS data (Figures 25 and 26) show exactly the same patterns as the other two datasets though a higher number of species have low abundance, reflecting the very heterogeneous nature of the vegetation categories in this dataset. The low median value of the *c* parameter, 0.37, is probably also a reflection of this. With this higher number of low abundance species it becomes clearer that there is an anomalous group of species with high *c* and *d* values. It is likely that the parameter values of this group arise from problems with curve fitting due to the lack of information. This problem would be more acute for the CS dataset with its smaller number of nests.

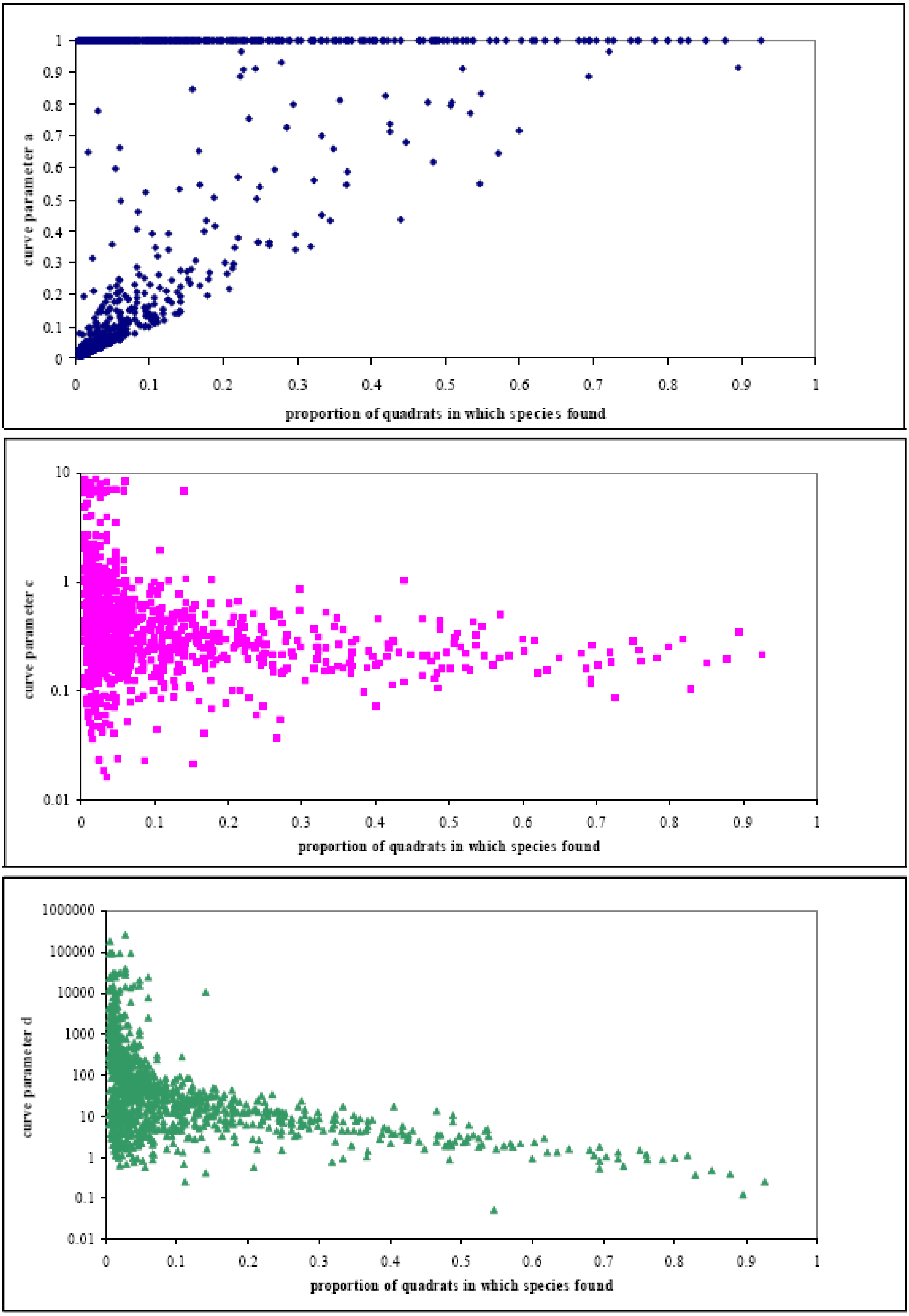


Figure 25 CS data – Parameters values and species frequency

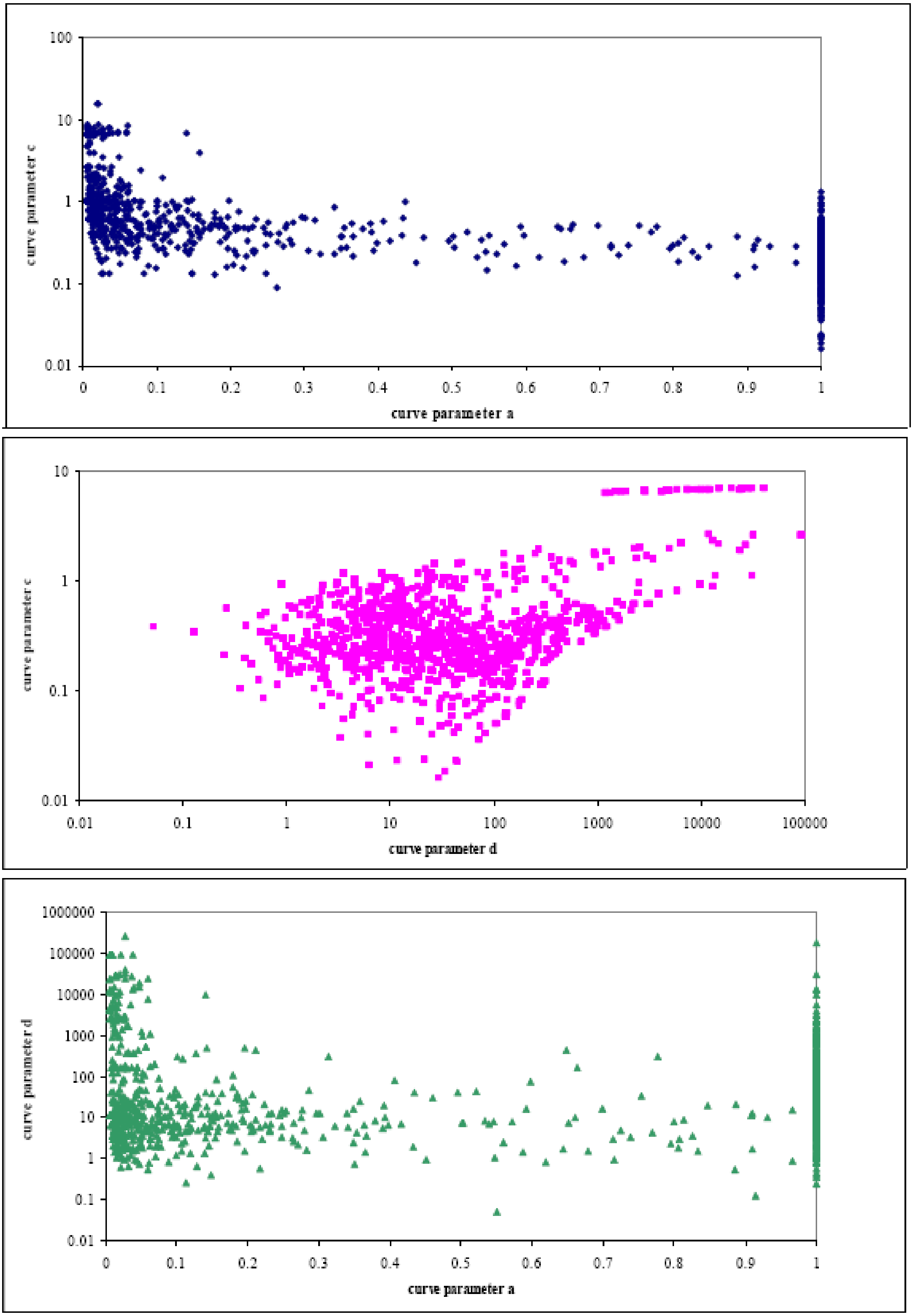


Figure 26 CS data – Relationship between parameters values

Overview

5.10 In Section 4 it was shown that the same models were appropriate whether the data represented homogeneous vegetation, such as the individual NE sites, or very heterogeneous vegetation such as that from the complete CS dataset. Examination of the variation in parameter values further emphasises this lack of difference between the datasets since the same patterns are found in each.

5.11 Two main findings result from the examination of the parameter values from the fitted logistic *acd* model. Firstly the fitted a parameters are essentially divided into those having values of either one or the maximum observed frequency. Secondly there is a relationship between the *c* and *d* parameters that results in species being distributed over a triangular region of the parameter space suggesting that high density species exhibit a different form of spatial patterning.

6 Effect of quadrat size on community level analysis

6.1 The effect of quadrat size on analysis has frequently been discussed in the literature but has rarely been quantified. In this study ordination techniques are used to examine size effects. Such techniques are commonly used in vegetation studies and are particularly useful in that although quantitative results are obtained, and can be dealt with statistically, they can also be presented visually giving an immediate impression of results without the need for analytical interpretation required by more algebraic methods.

6.2 Ordination diagrams are used here to show how samples move about the ordination space as different levels of nesting are used. They can also be used to show differences between the true dataset at a given nest size and the equivalent dataset produced by standardisation from a smaller (or larger) nest and this facility will be utilised in the next section of the report.

Natural England data

6.3 Figure 27 shows the first two dimensions of the ordinations obtained from the Natural England data by performing separate analyses at each of the nest sizes. The most striking aspect of these diagrams is the similarity of the results. There is a clear distinction between sites from the three vegetation types, and the relative positioning of the groups of sites is the same for each nest size. Within each NVC type there is some difference in the positioning of points, with a suggestion of greater separation, and hence greater discrimination, as the nest size increases.

6.4 Figure 28 shows ordination diagrams of just the MG3 sites. While there are some differences between the individual plots, the relative positioning of points remains remarkably stable confirming that the apparent stability shown in Figure 27 is not just an effect of the strong differences between NVC types dominating the ordination at all nest sizes.

6.5 Figures 29 and 30 are the corresponding diagrams for ordination dimensions 3 and 4. Similar comments are appropriate here. The results of analysis are essentially the same regardless of nest size though there is somewhat more variation between plots than in Figures 27 and 28.

6.6 To obtain more detailed information about the differences between nest sizes a combined analysis was obtained. All nest sizes were entered into a single analysis in which the information at each nest size is treated as if it came from a separate quadrat. Each site is therefore represented in the output by a set of points, one for each nest size. Connecting the points representing a site, in order of nest size, shows the differences between nest sizes and provides a basis for assessing cumulative or consistent differences. Figure 31 shows the ordination diagram for dimensions 1 and 2 and Figure 32 the diagram for dimensions 3 and 4. The results are striking with sites represented by lines of points streaming out from a central area. The smallest nest size is at the centre of each diagram and the largest at the periphery. It is clear, therefore, that the nested data does contain information about the scale of measurement but that this information is not apparent in ordinations from individual nest sizes (Figures 27 to 30) because of the scaling procedures used within the ordination process.

6.7 Figure 33 shows a combined analysis for just the MG3 sites. The same effect is apparent. This figure also suggests that such plots may have a useful analytical function in their own right, enabling the effect of varying quadrat size to be examined in detail. At the smallest nest size, for example, the two Borrowbeck sites and the Bowber Head site are approximately equidistant in the plot. As nest size increases, however, the two Borrowbeck sites become closer together and diverge from the Bowber Head site. Interpretation of this effect is, however, difficult. It could be argued that increasing nest size has brought out the similarity of the two Borrowbeck sites.

Alternatively it could be that the larger nest size has masked essential differences that are apparent at the smaller nest size.

6.8 Figures 34 and 35 illustrate the effects on analysis of mixing nest sizes. In each figure nest size is varied across vegetation types and/or sites. In Figure 34 data for the MG3 and CG2 sites was taken from the smallest nest size while data for the CG5 sites was taken from the largest. The result is that the MG3 and CG2 sites form compact clusters close to each other while the CG5 sites are much more widely spread and far from the other two vegetation types. The impression is very different from that given by any of the diagrams in Figure 27.

6.9 In Figure 35 a different choice of nest sizes was made. The sites were ordered alphabetically and the smallest nest size used for the first eight sites. The largest nest size was used for the remaining seven sites. The result is once again very different from any of the individual plots in Figure 27. It could almost be argued that the sites fall into the classic “horseshoe” shape frequently found in ordination diagrams. Furthermore the clear separation between NVC types has broken down.

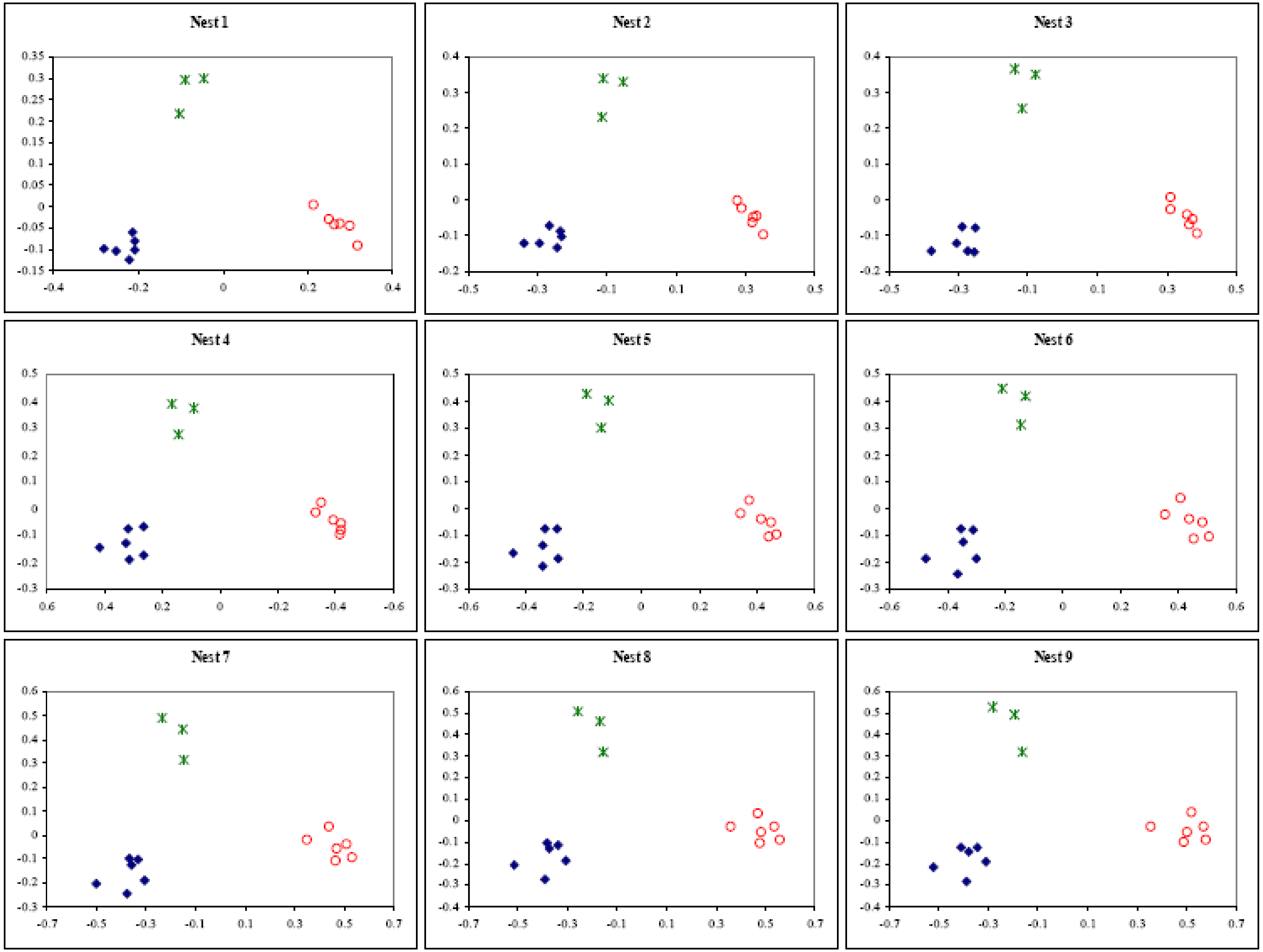


Figure 27 Natural England Data - Correspondence analysis of sites by nest size, dimensions 1 and 2

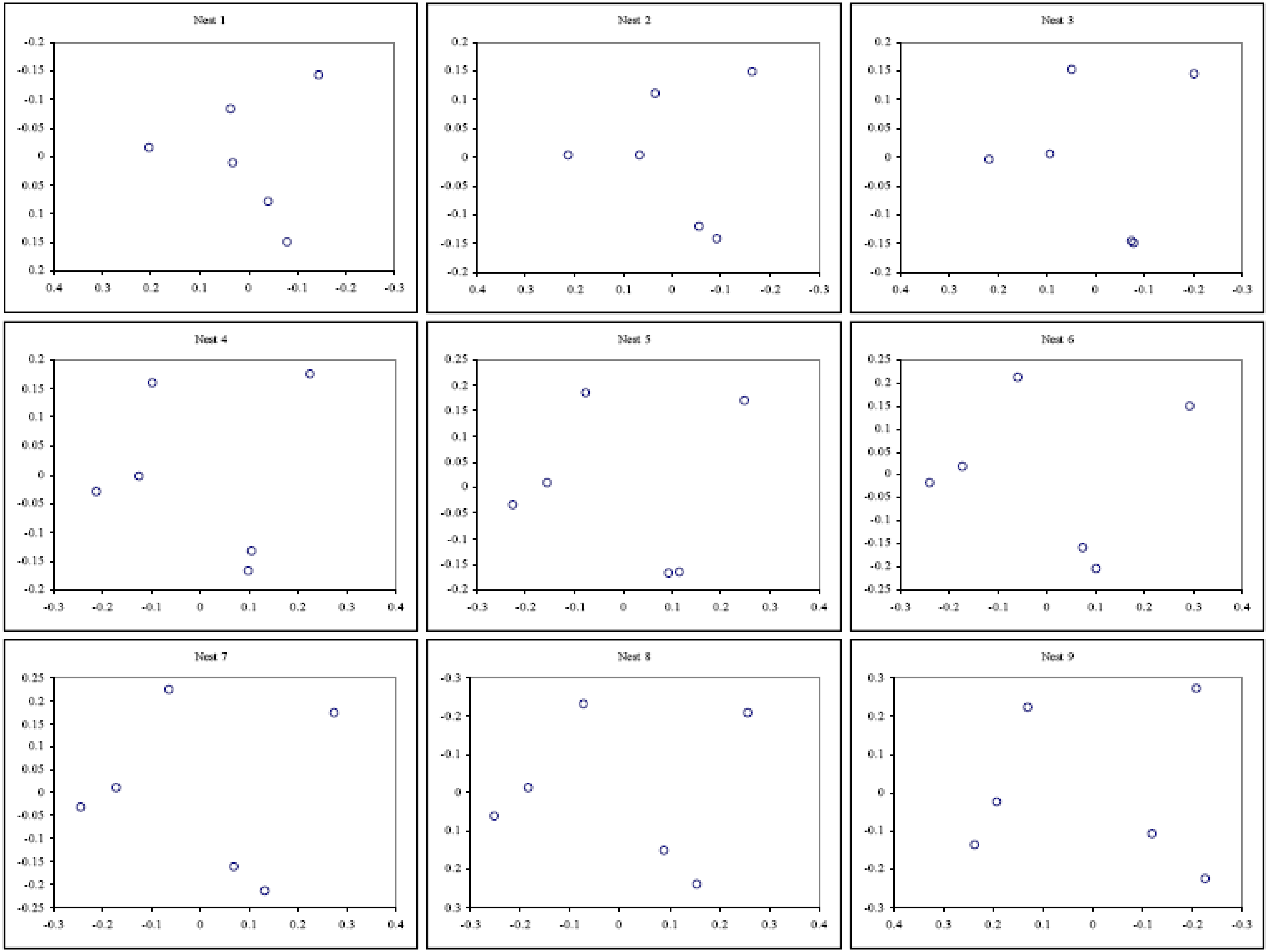


Figure 28 Natural England MG3 sites - Correspondence analysis of sites by nest size, dimensions 1 and 2

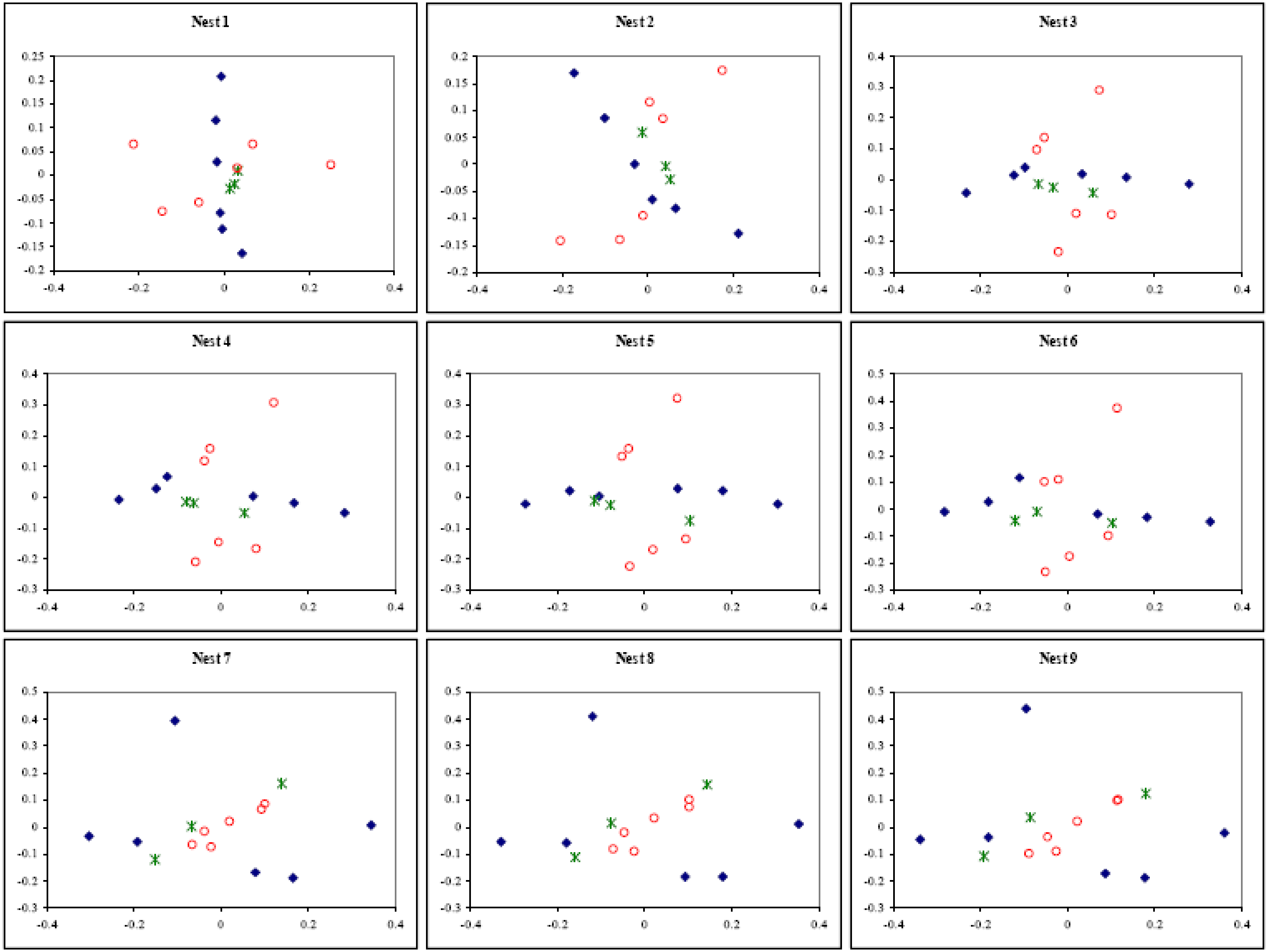


Figure 29 Natural England Data - Correspondence analysis of sites by nest size, dimensions 3 and 4

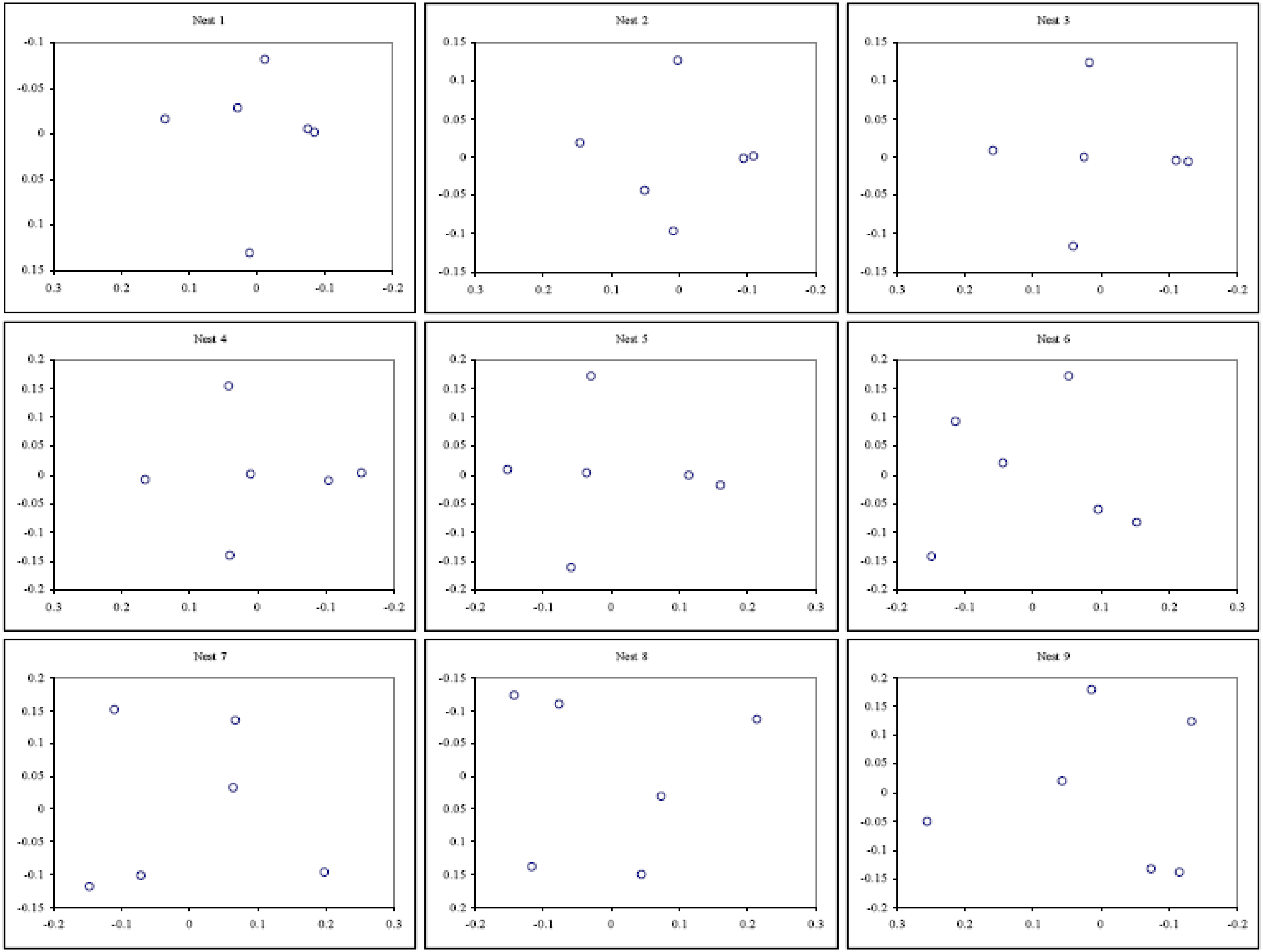


Figure 30 Natural England MG3 sites - Correspondence analysis of sites by nest size, dimensions 3 and 4

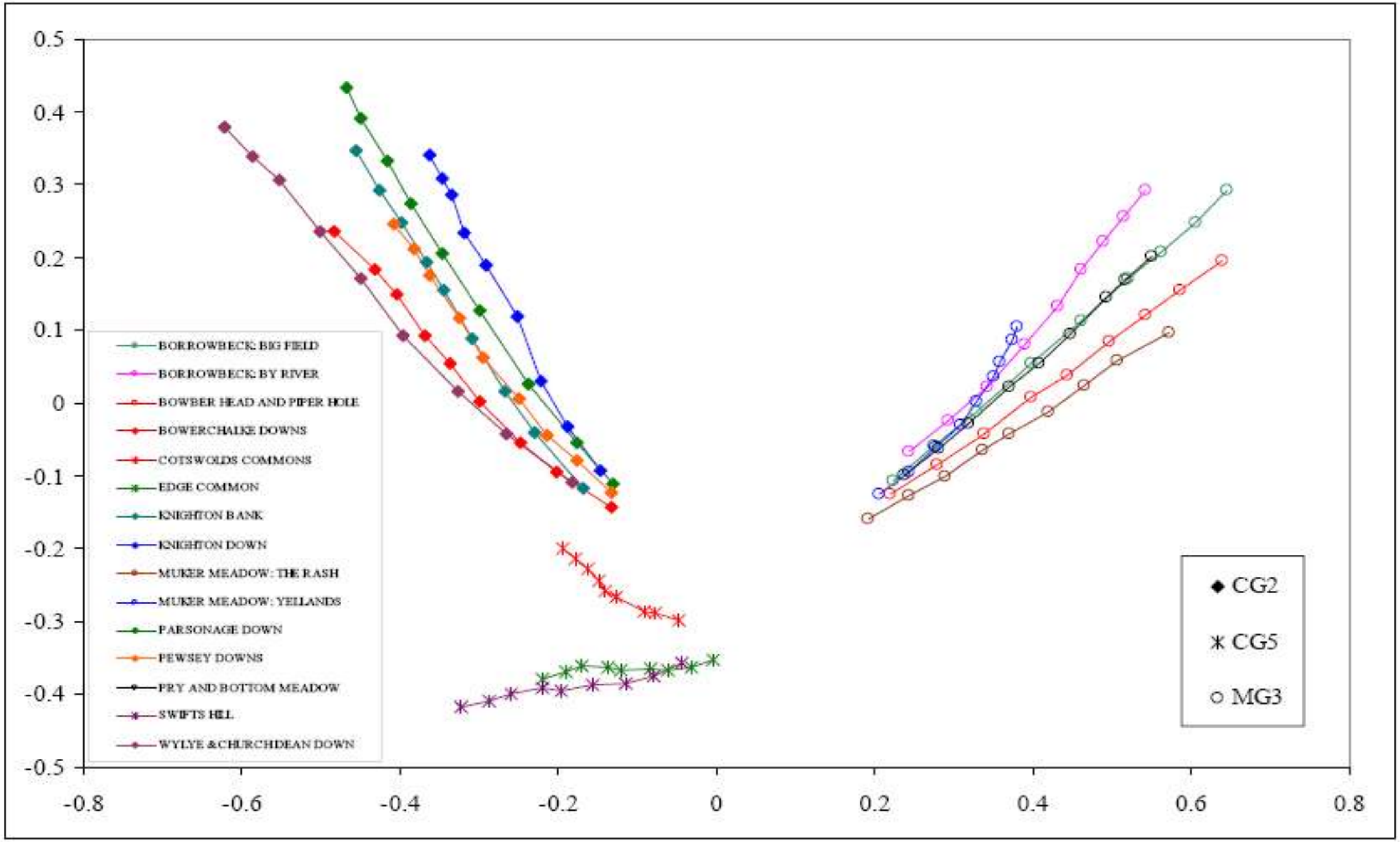


Figure 31 Natural England data Correspondence analysis dimensions 1 and 2 - nests entered as separate points

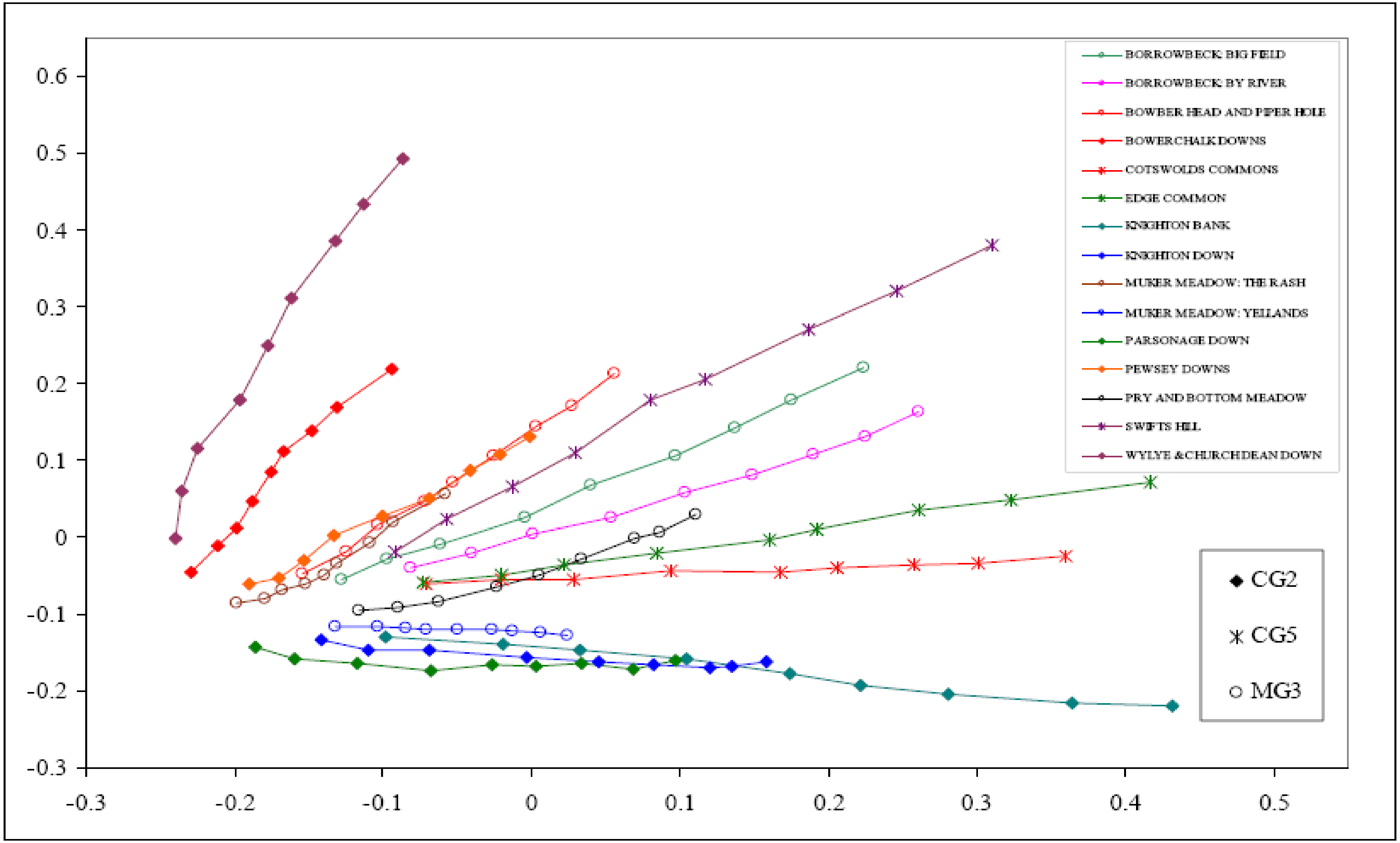


Figure 32 Natural England data Correspondence analysis dimensions 3 and 4 - nests entered as separate points

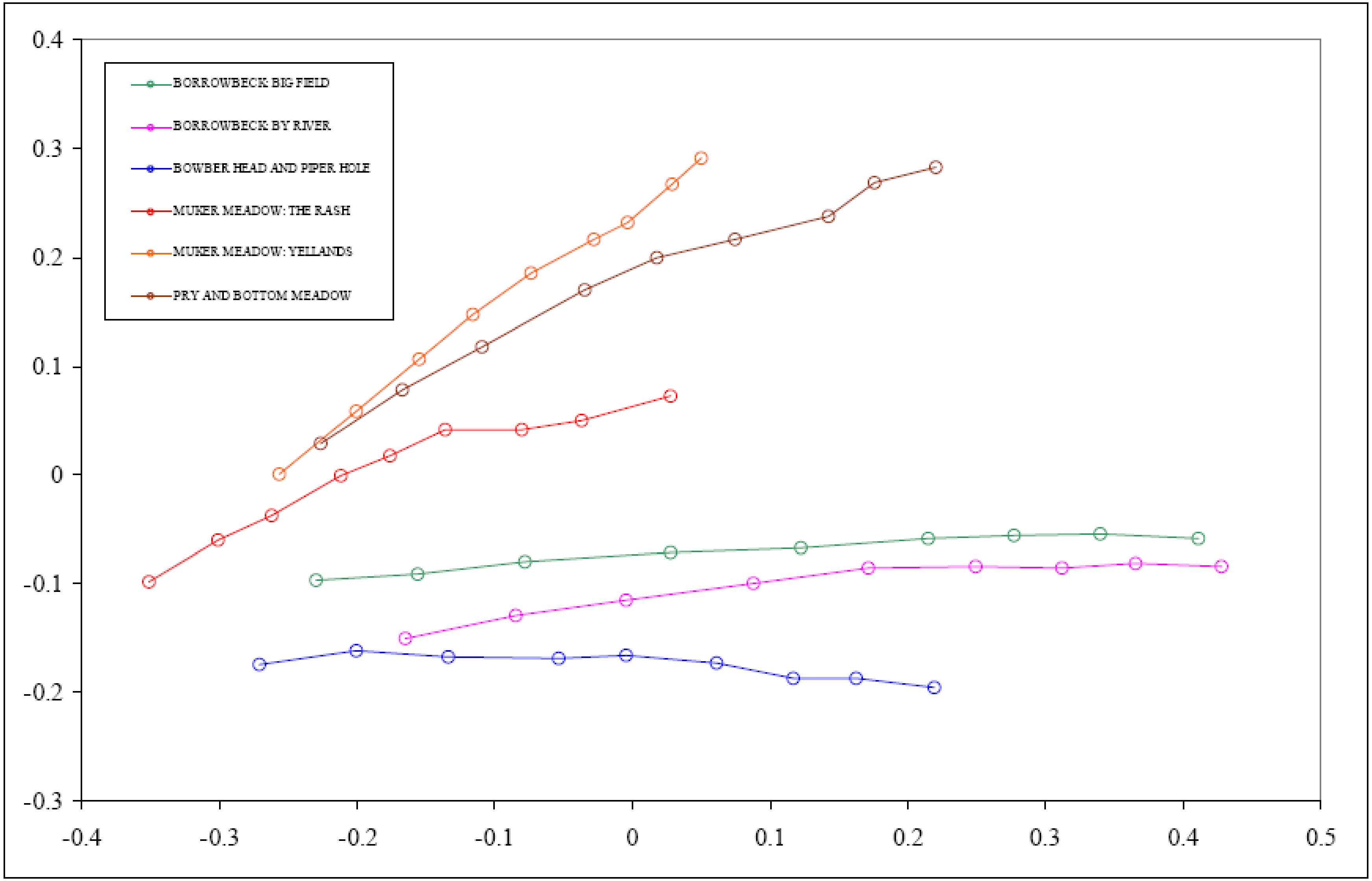


Figure 33 Natural England MG3 sites Correspondence analysis - nests entered as separate points

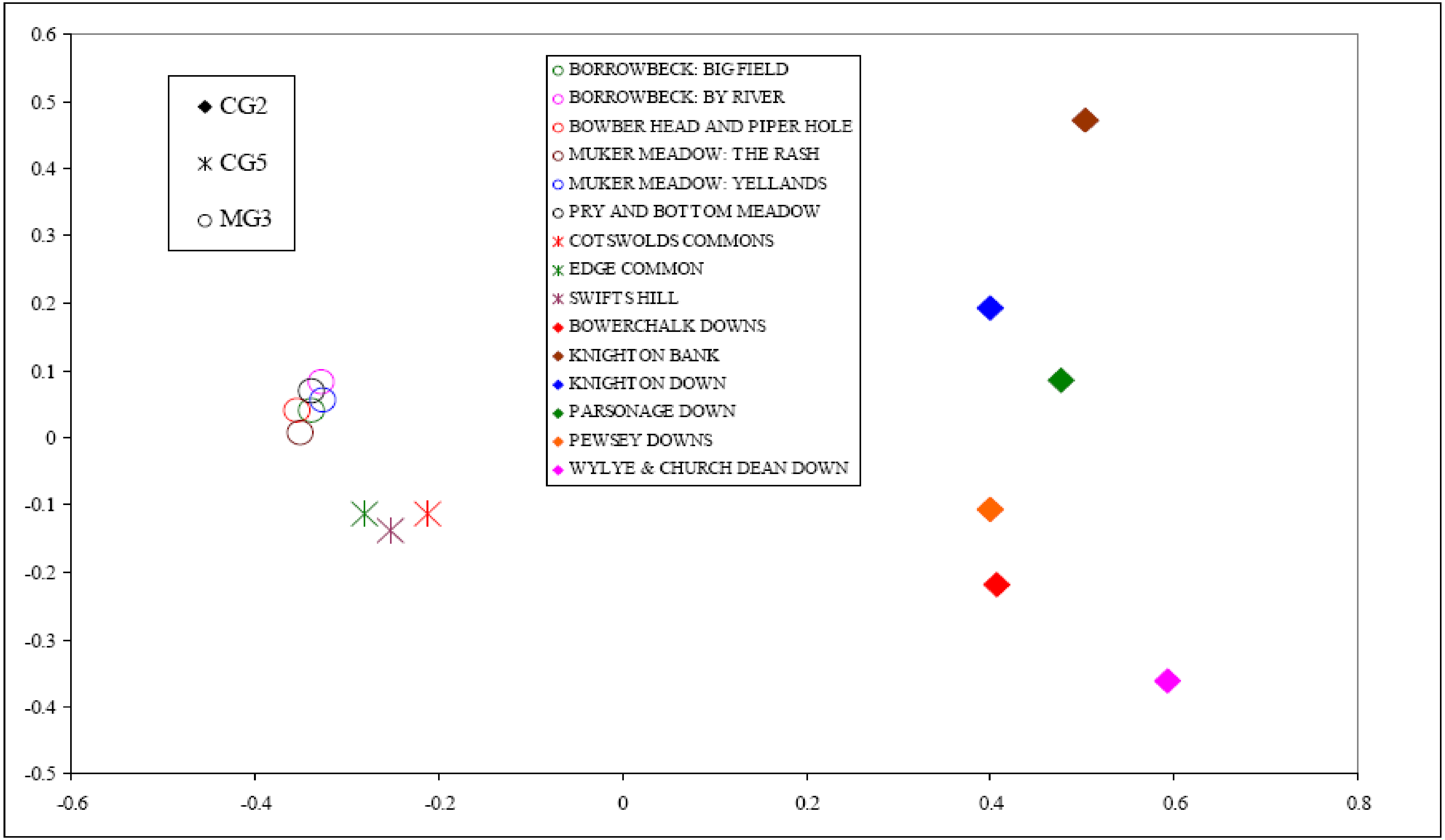


Figure 34 Natural England data Correspondence analysis MG3, CG2 sites with smallest nests, CG5 sites with largest nests

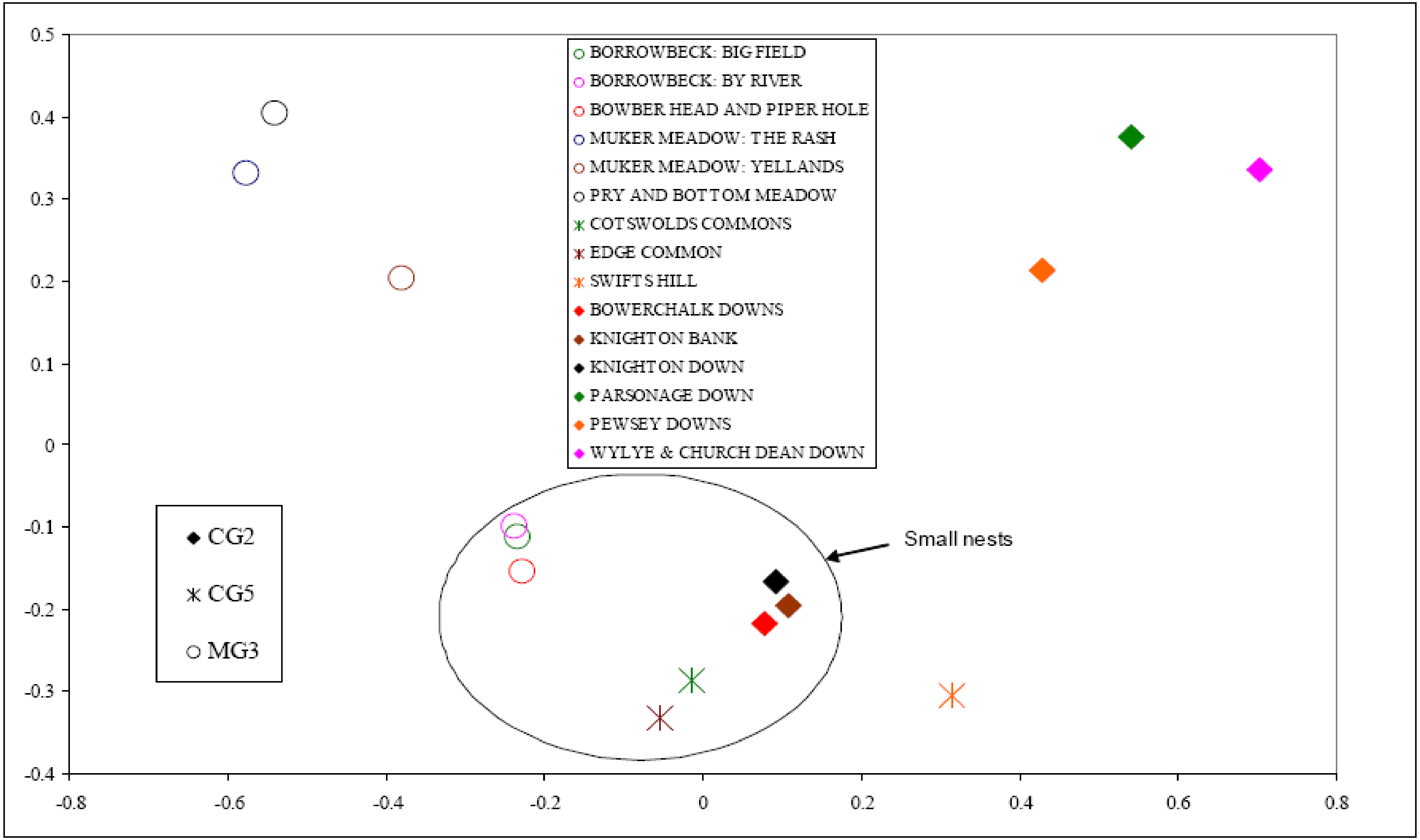


Figure 35 Natural England data Correspondence analysis - nest size varied across NVC category

ADAS ESA data

6.10 Figures 36 and 37 show the first four dimensions of the individual correspondence analyses of the sixteen nest sizes of the ESA data. As with the NE data the results are very similar for all nest sizes, though again a small degree of gradual change can be seen when moving from the smallest to largest nests. There is an obvious outlier on the first axis, the South Wessex Downs, and on the third and/or fourth axis, the Lake District. Both are picked out at all nest sizes.

6.11 Figures 38 and 40 show the first four correspondence analysis axes for the combined analysis of all nest sizes for the ESA data. The same spreading out from a central point is seen as with the NE data. Also obvious are the two outliers seen in the individual plots. Figures 39 and 41 show expanded views of the same information with the outlying points removed, to enable details to be picked out more easily. The impression given by Figures 38 and 39 is that the sites move apart in two opposite directions. Examination of the activities listed in Table 2 show that those sites apparently moving in a south-east direction with increasing nest size in these two figures are the moorland/heathland sites while the grassland sites move to the north-west. The first South West Peak site, which is a mixture of moorland and grassland, shows least movement, remaining at the centre of the plot. Thus, as with the NE data, these mixed nest plots appear to have a useful analytical capability.

6.12 Figure 42 illustrates the effect of mixing nest sizes. In this example the largest nest size was used for the first six sites, when they are arranged alphabetically, an intermediate nest size for the next six and the smallest nest size for the last six. Once again the structure of the plot has little resemblance to the plots in Figure 36 and in particular the South Wessex Downs, a very clear outlier in plots 35 and 37, is no longer presented as such.

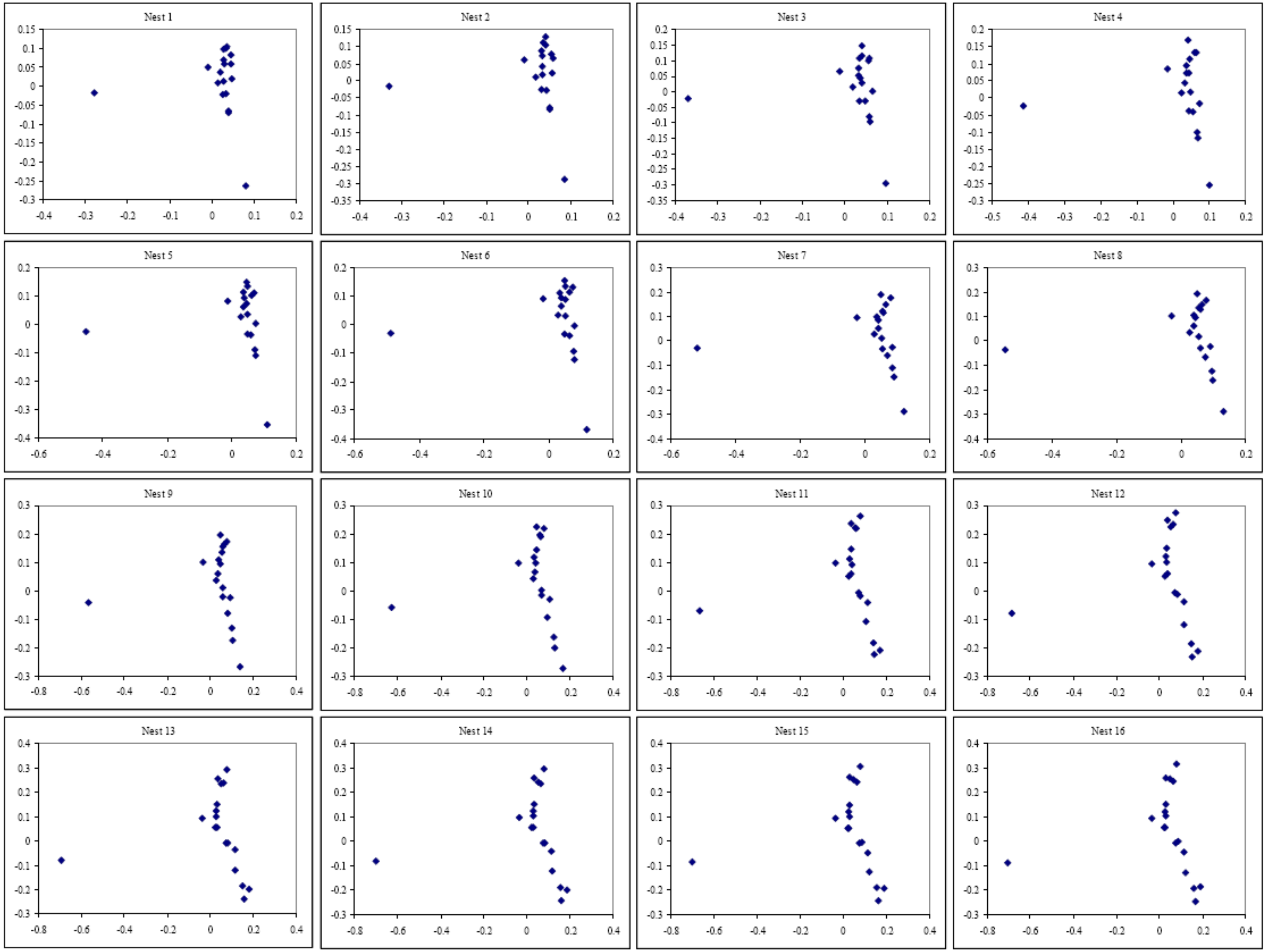


Figure 36 ADAS ESA data - Correspondence analysis of sites by nest size, dimensions 1 and 2

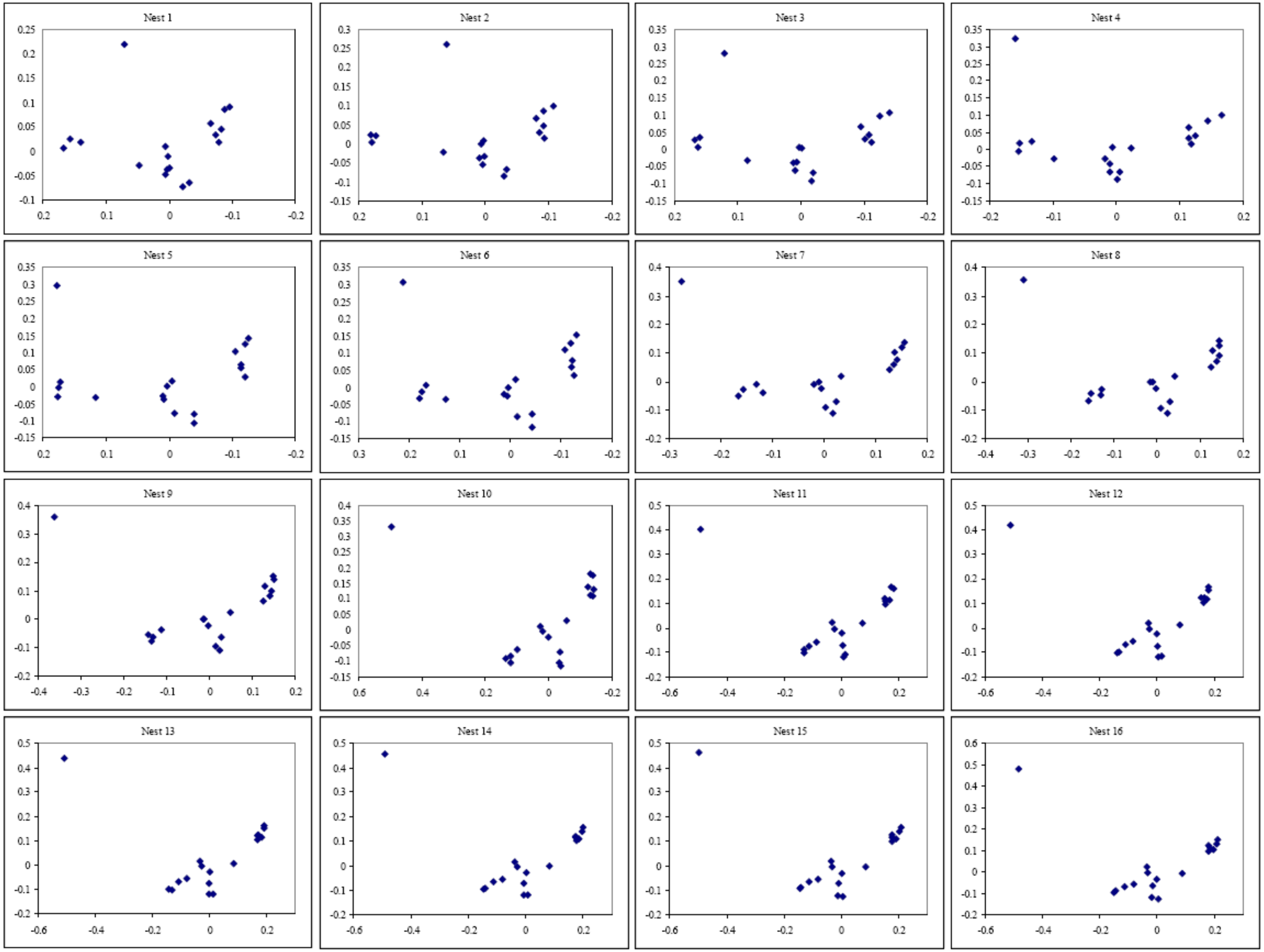


Figure 37 ADAS ESA data - Correspondence analysis of sites by nest size, dimensions 3 and 4

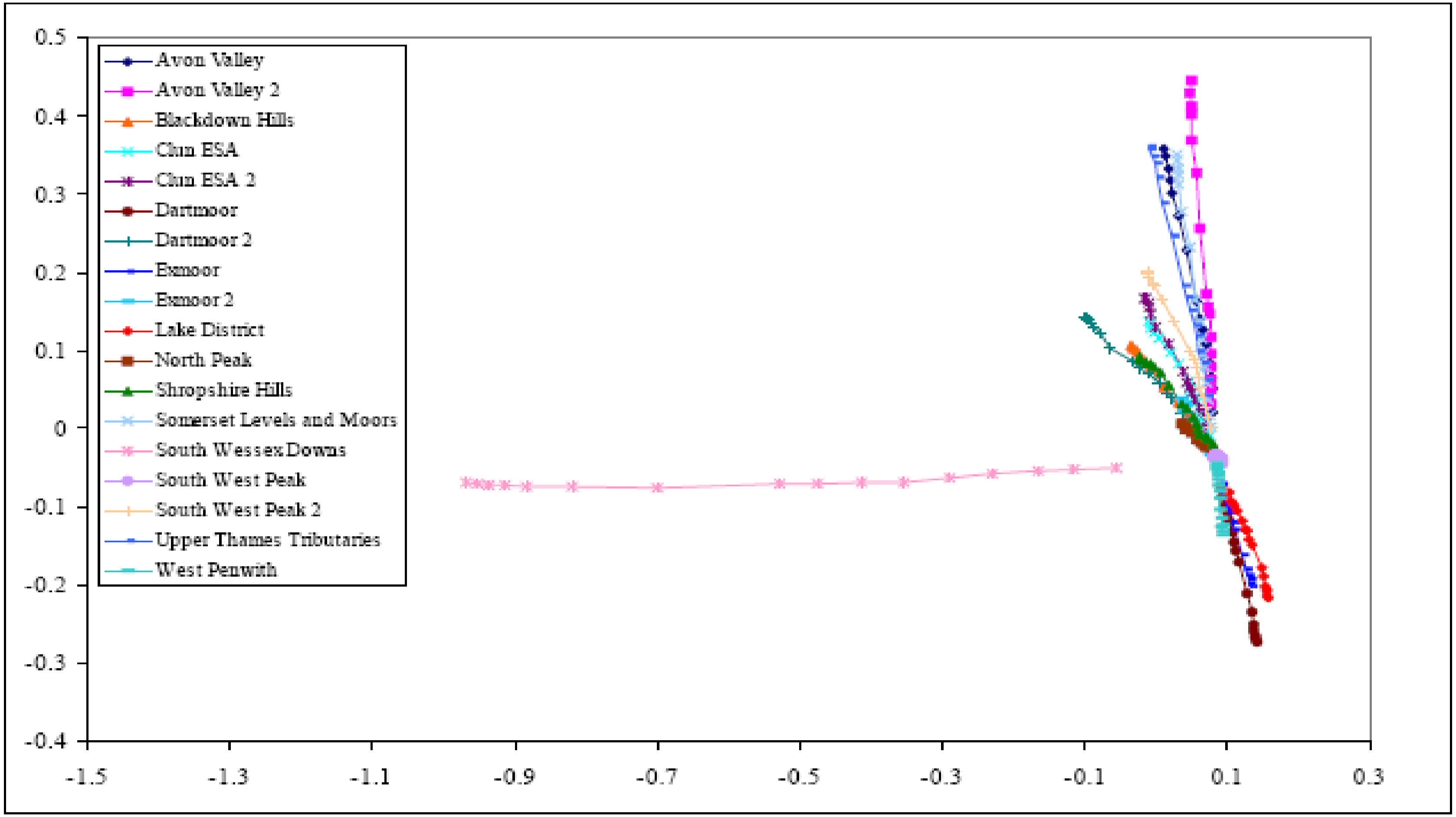


Figure 38 ADAS ESA data Correspondence analysis dimensions 1 and 2 - nests entered as separate points

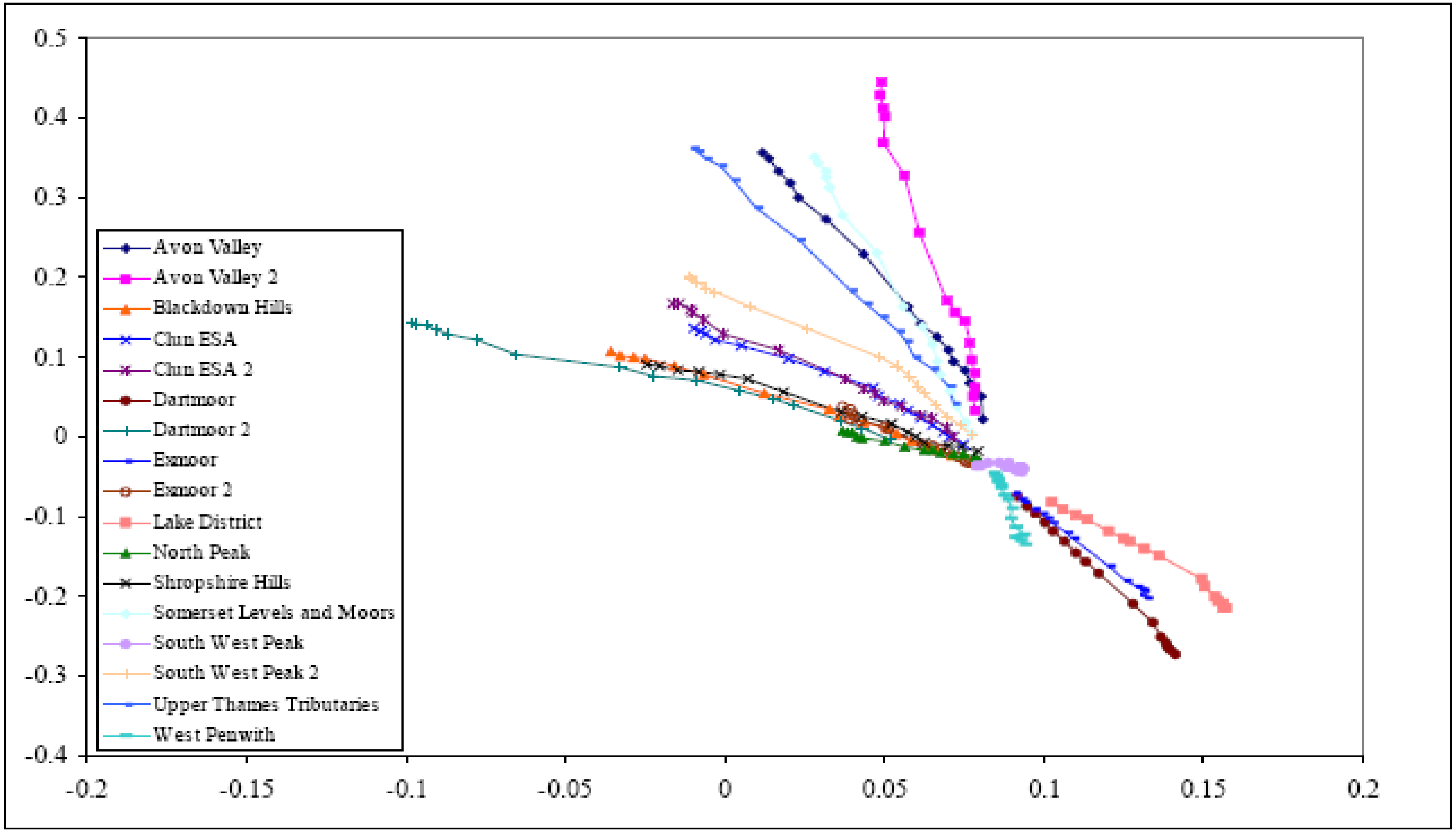


Figure 39 ADAS ESA data Correspondence analysis dimensions 1 and 2 - nests entered as separate points (South Wessex Downs omitted)

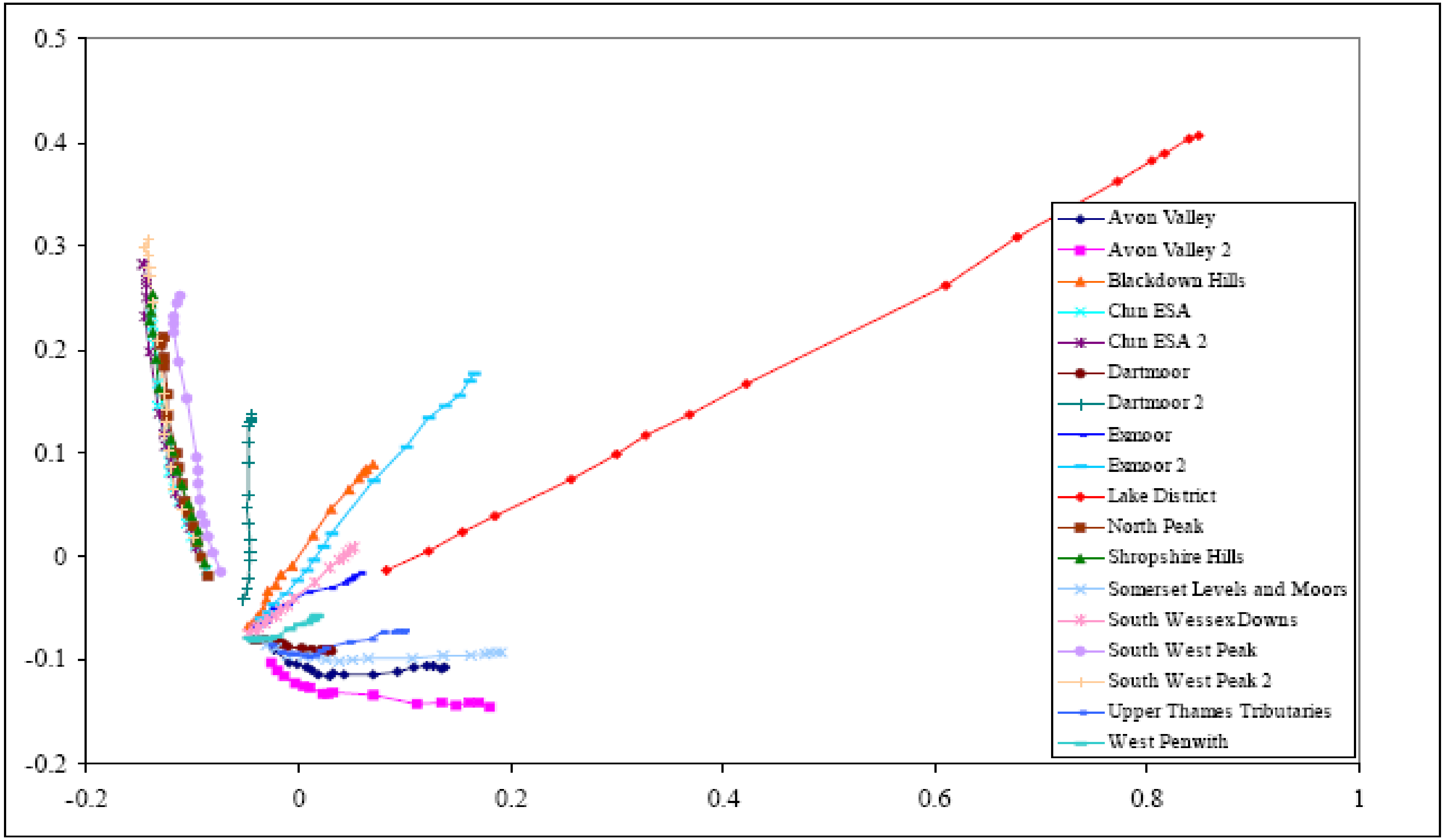


Figure 40 ADAS ESA data Correspondence analysis dimensions 3 and 4 - nests entered as separate points

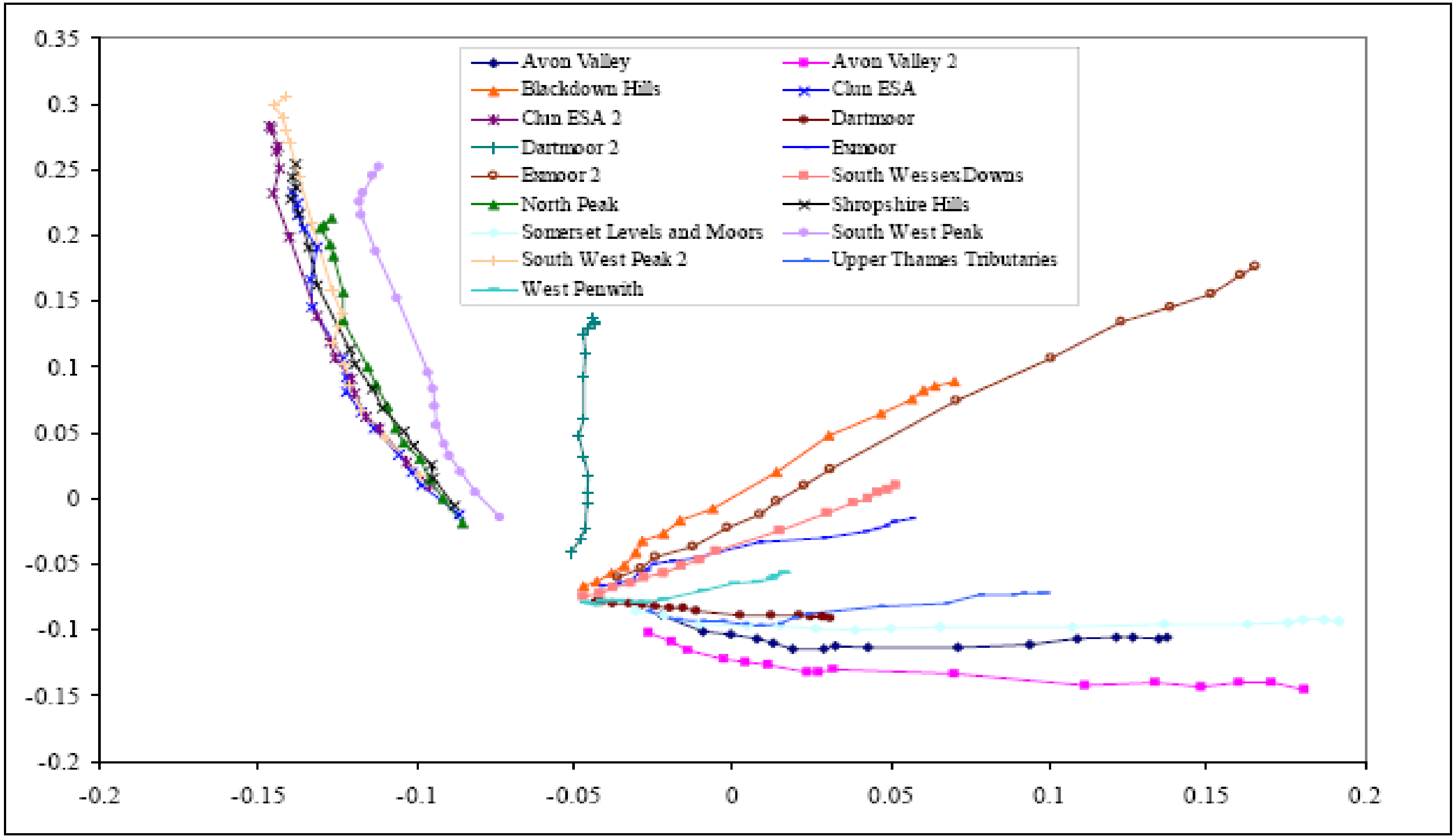


Figure 41 ADAS ESA data Correspondence analysis dimensions 3 and 4 - nests entered as separate points (Lake District omitted)

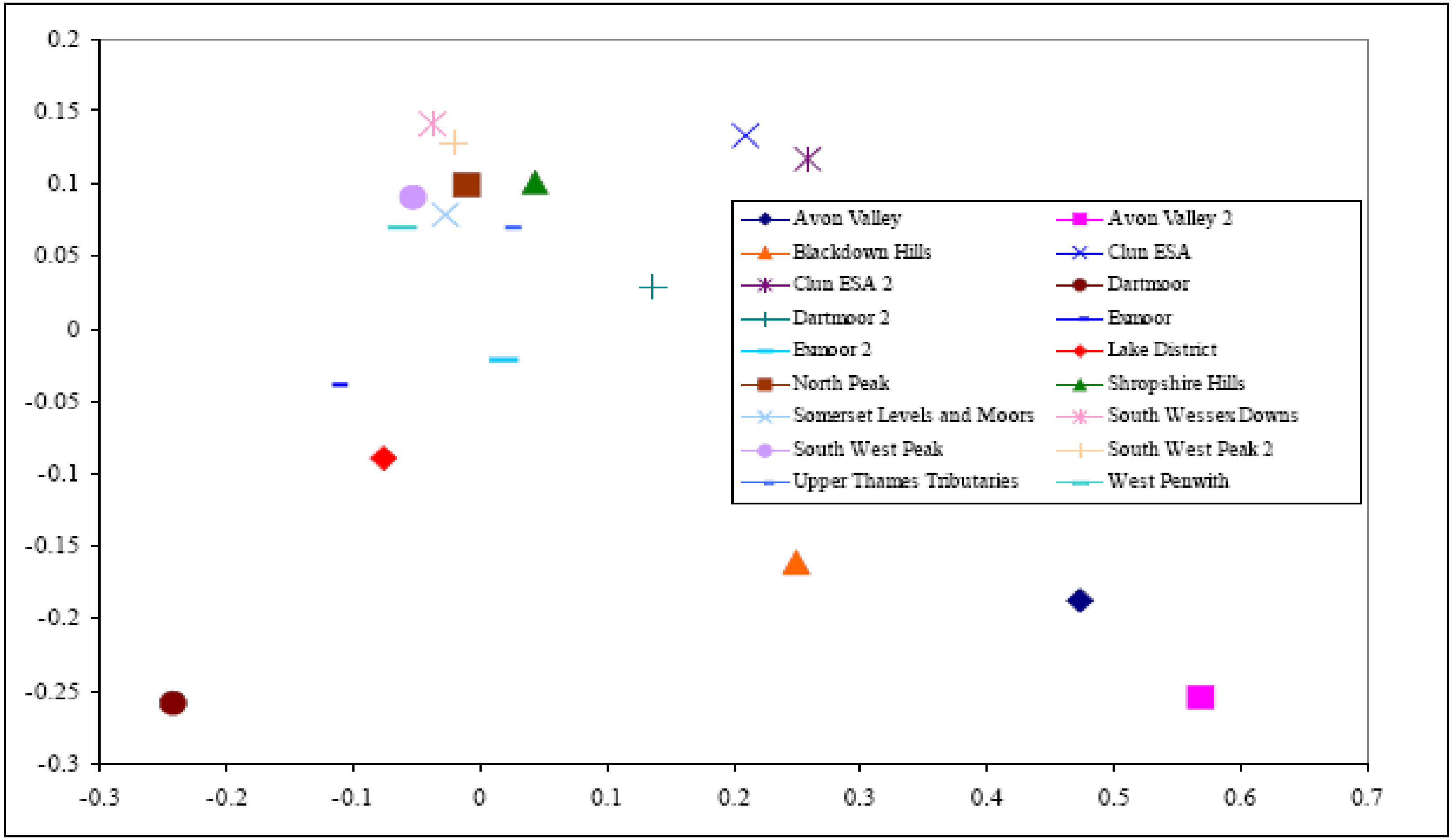


Figure 42 ADAS ESA data Correspondence analysis dimensions 1 and 2 - nest size varied with site

6.13 Figure 43 shows plots of the first four correspondence analysis axes for the five individual nest sizes of the countryside survey data. Even more than the previous two datasets these plots are remarkably similar, with very little variation across nest sizes. Aggregate class 5, Lowland Wooded, appears as an outlier on the third axis.

6.14 Figures 44 and 45 show the combined plot for all nest sizes. The same factors are apparent as for the other two datasets, with the points representing individual aggregate classes radiating out from a central point in an almost linear fashion. Unlike previous combined plots the lines representing classes start further from the central point and are more widely separated, which may reflect the larger starting nest size.

6.15 If the plots on the left hand side of Figure 43 are divided horizontally so as to separate the top four points from the bottom five, then it is clear that the topmost points show much less variability than the remainder. This feature was used to examine the effect of mixing nest sizes in this dataset. Data from the largest nest size for aggregate classes AG0, AG1, AG2 and AG5, those represented by the top four points, were combined with data for the smallest nest size of the remaining classes. Figures 46 and 47 show the results. The main effect is that aggregate class 5, which was an outlier on the third correspondence analysis axis for all nests sizes (Figure 43) is now an outlier on the first dimension (Figure 46) and is centrally placed on dimensions three and four (Figure 47). Thus the effect of mixing nest sizes in this case has been to exaggerate the importance of a relatively minor feature of the data.

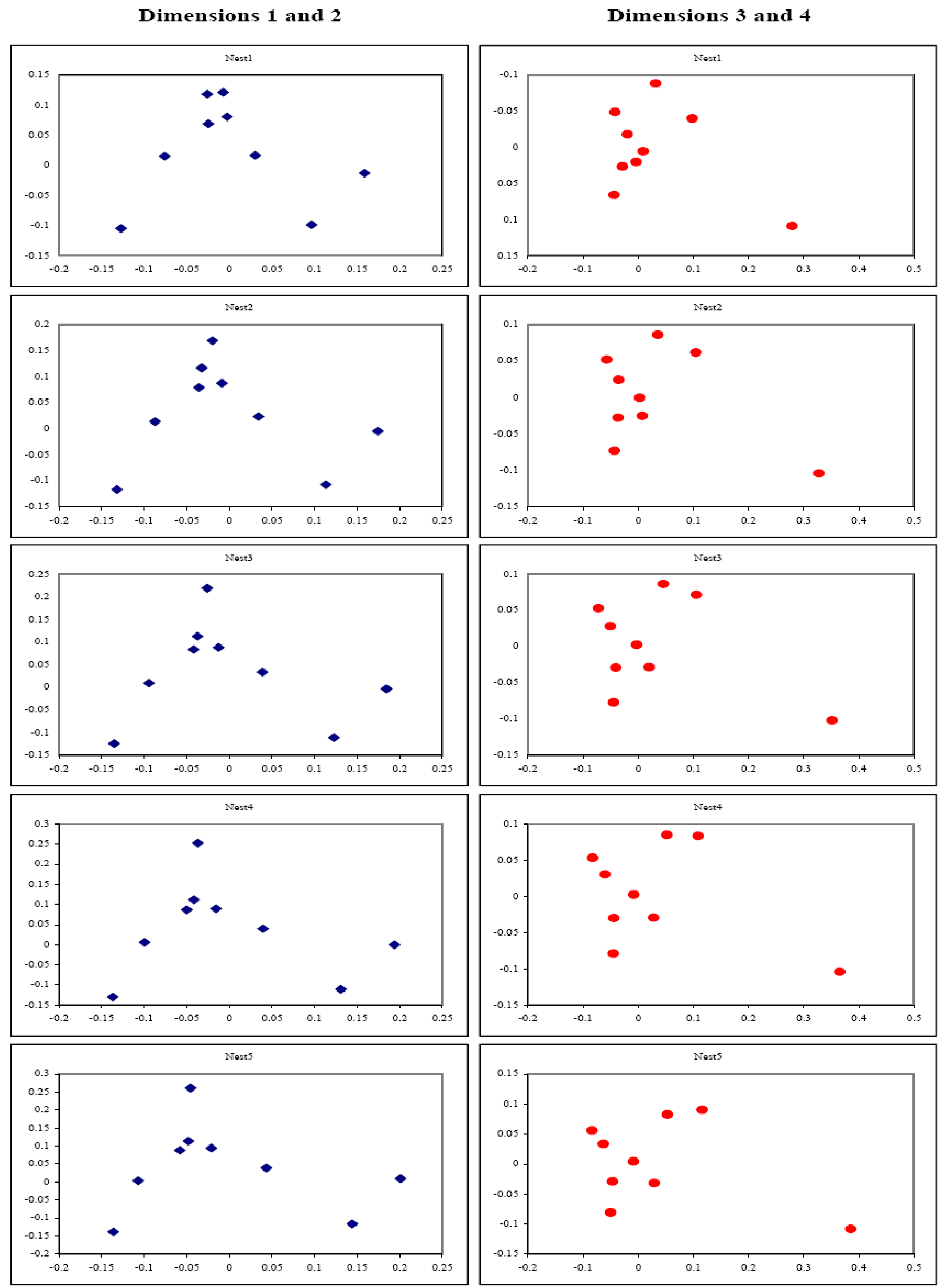


Figure 43 Countryside Survey Correspondence analysis by CVS Aggregate Class

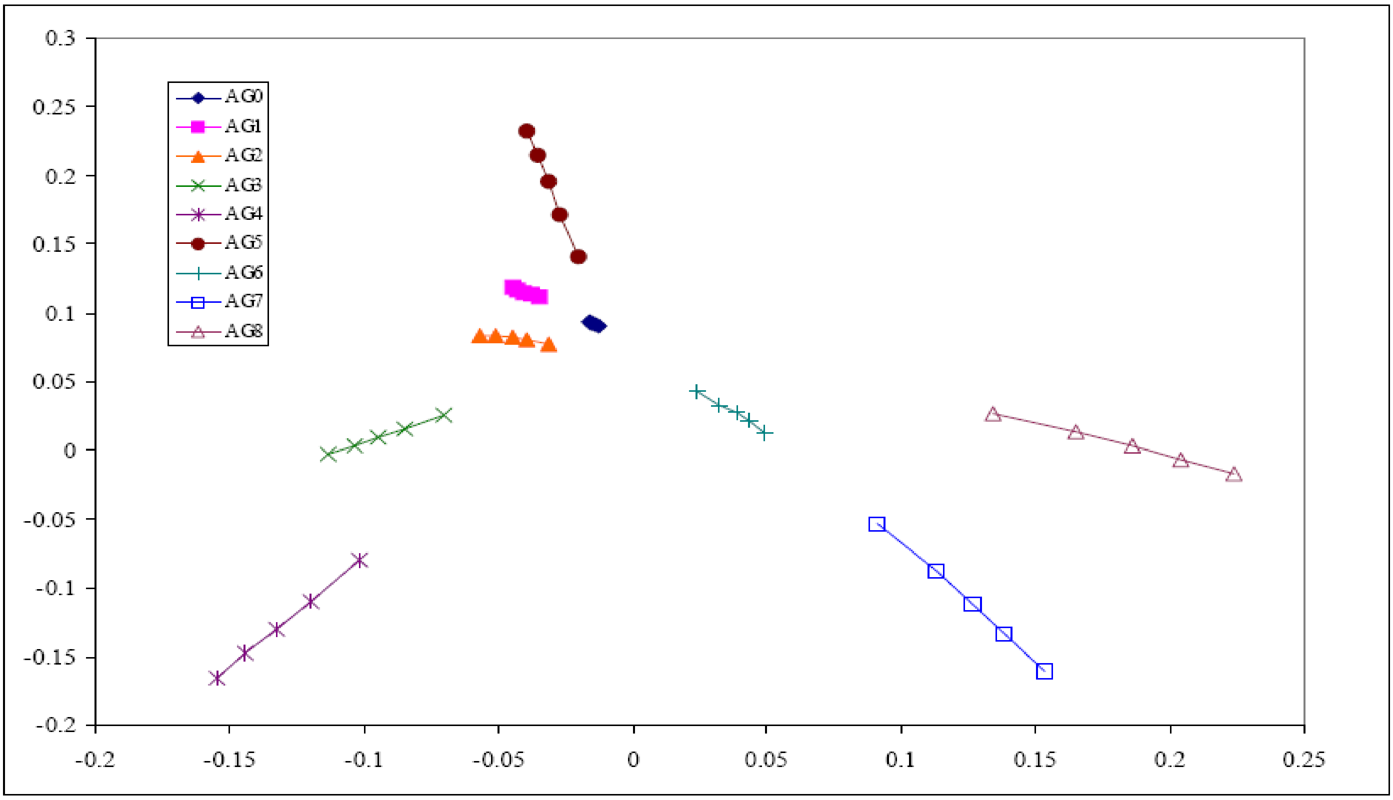


Figure 44 CS data Correspondence analysis dimensions 1 and 2 - nests entered as separate points

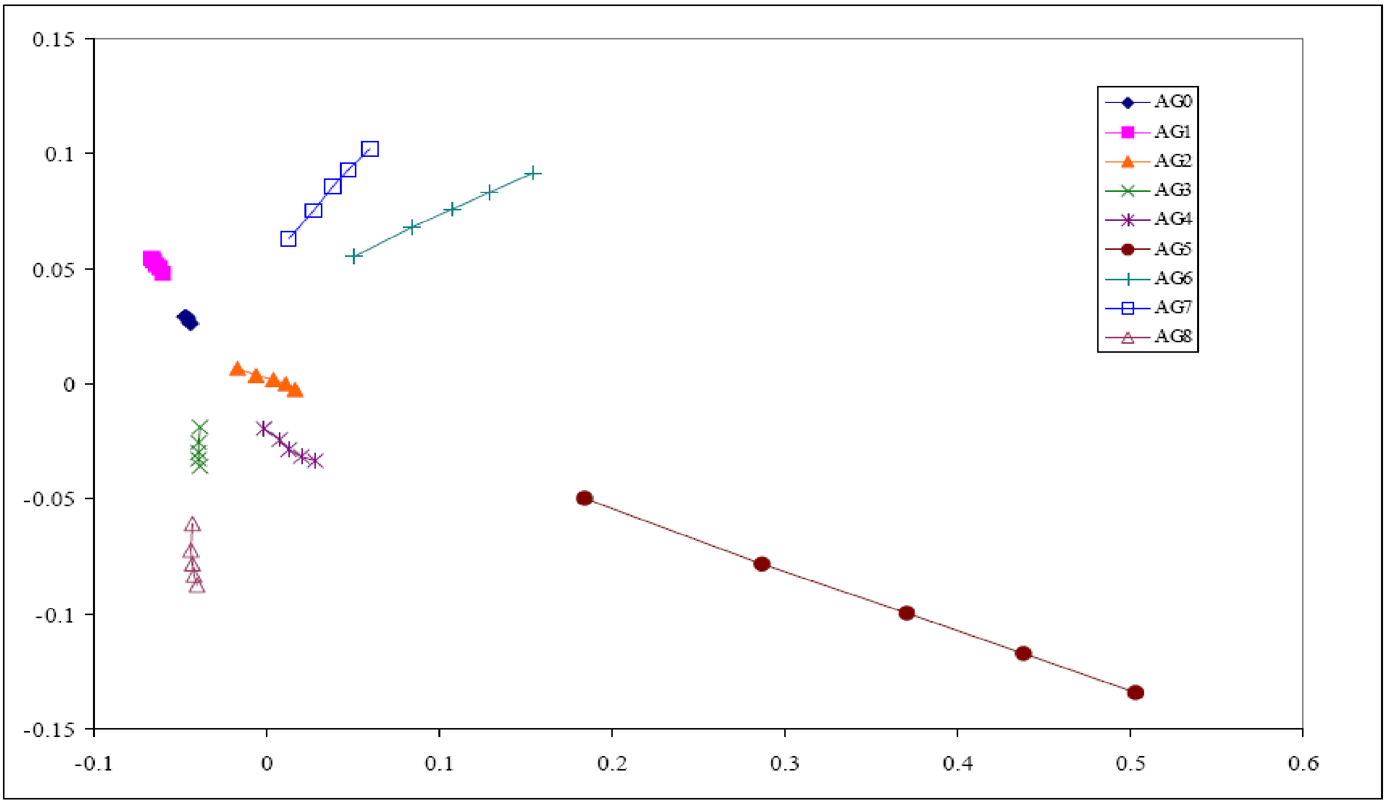


Figure 45 CS data Correspondence analysis dimensions 3 and 4 - nests entered as separate points

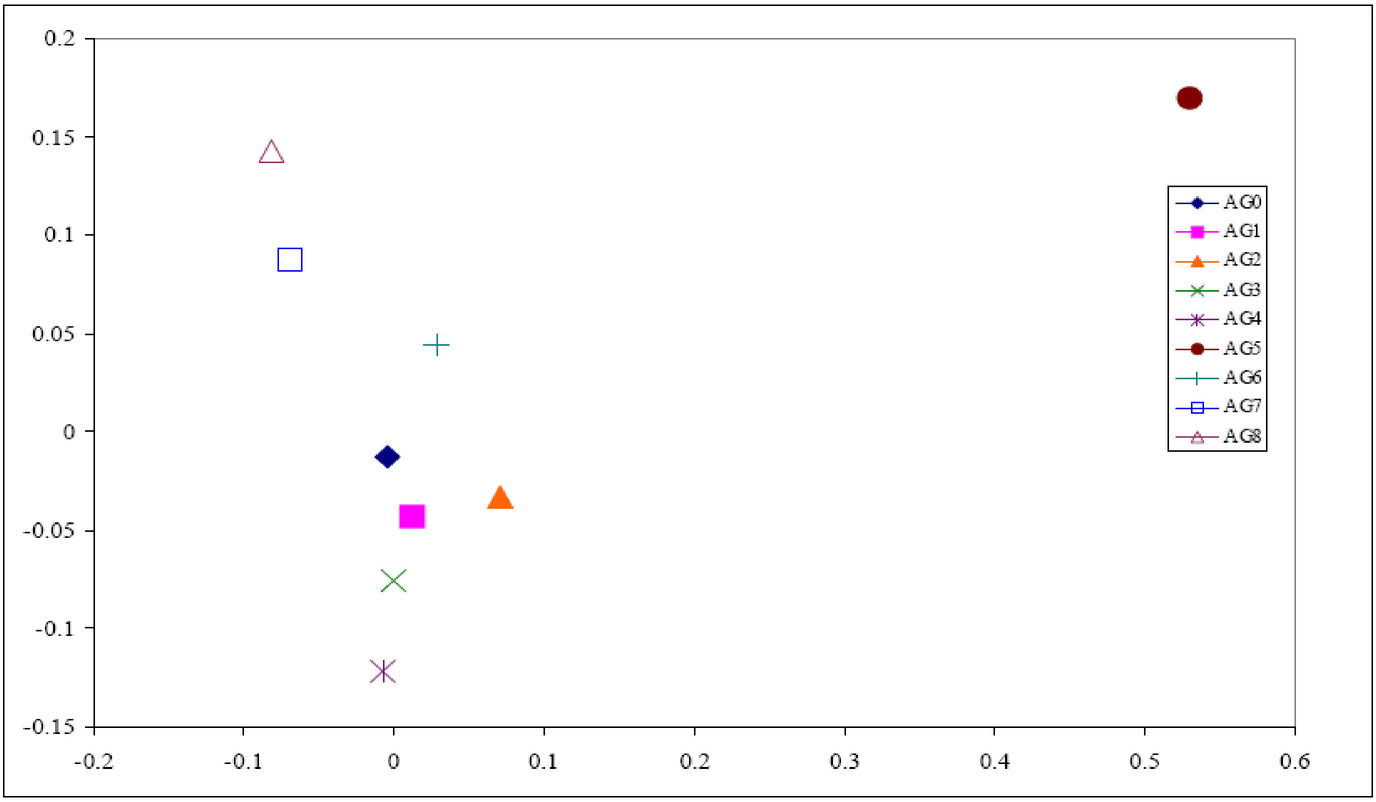


Figure 46 CS data Correspondence analysis dimensions 1 and 2 AG0, AG1, AG2 AG5 nest size 1: AG3, AG4, AG6, AG7, AG8 nest size 5

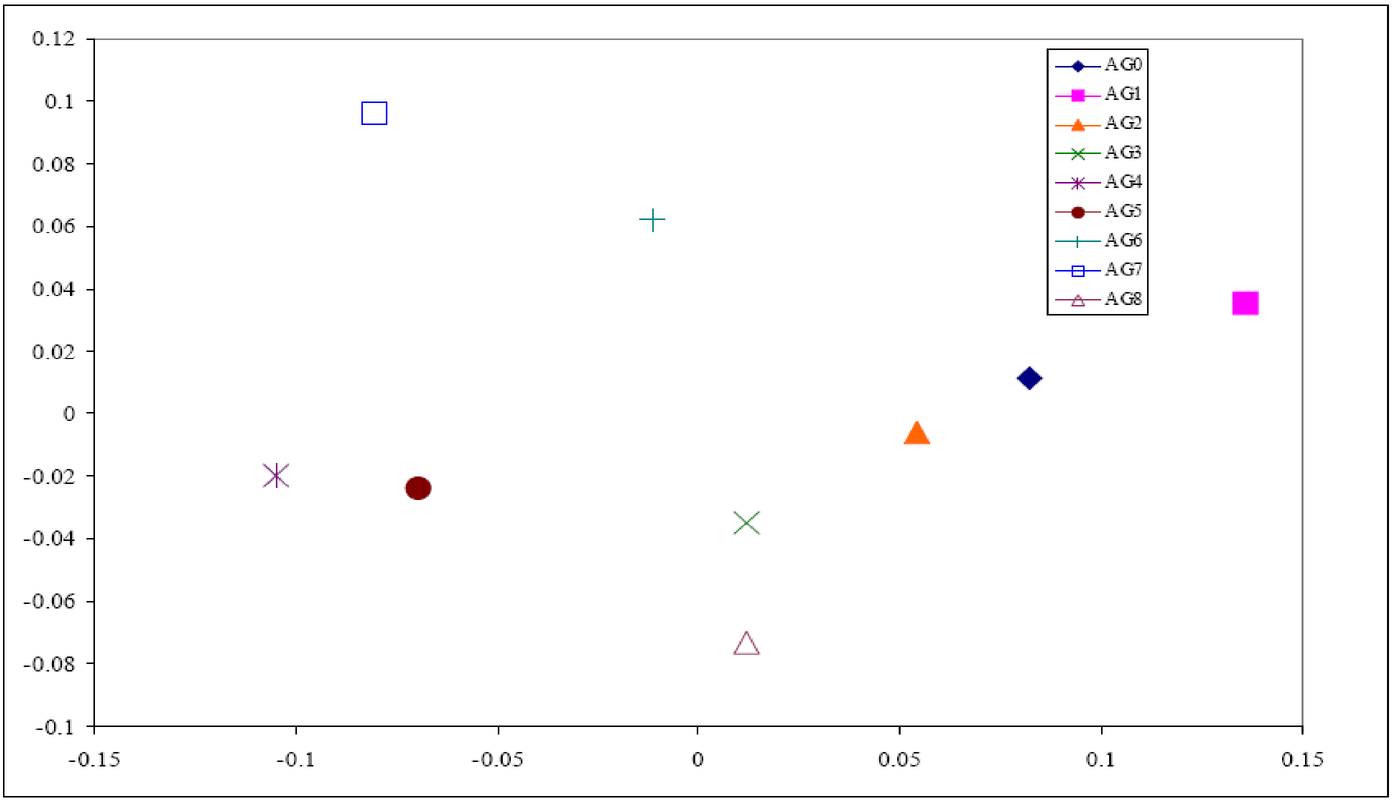


Figure 47 CS data Correspondence analysis dimensions 3 and 4 AG0, AG1, AG2 AG5 nest size 1: AG3, AG4, AG6, AG7, AG8 nest size 5

Overview

6.16 The three datasets used in this study are extremely disparate in terms of heterogeneity of vegetation and range of nest sizes. The results presented in this section are, in contrast, remarkable uniform. Firstly, for each dataset, the ordinations obtained from each nest size are very similar suggesting that nest size is not crucial to analysis and that the same information and conclusions will be obtained regardless of which nest size is used. Secondly, datasets compiled from mixed quadrat sizes can give very different results from datasets in which all quadrats are the same size. In the mixed situation quadrat size tends to override the main axes of variation in the data. More remarkable however, as shown by the plots including all nest sizes, is that quadrat size doesn't just replace, or displace, one of the original axes but appears to affect all of them. In addition mixing quadrat sizes can suppress or exaggerate features of the data, such as the suppression of the outlying position of the South Wessex Downs in the ADAS data and the exaggeration of the difference of Aggregate Class 5 in the CS data. It should be noted, however, that the combined plots also indicate that the extent of the distortion will depend on how different the quadrat sizes are. In creating examples of mixed quadrat size datasets for illustrative purposes both extremes of size have been used in order to maximise the effects. Mixing adjacent nest sizes from any of the datasets does not produce such marked effects.

7 Standardisation

7.1 It was shown in the previous section that comparison of sites surveyed using different quadrat sizes can give a very distorted impression of site differences. If such comparisons need to be made then methods of correcting for the distortion may be essential if sensible interpretations of the data are to be obtained. In this section two alternative methods of correcting for differences in quadrat size are examined. The first is a simple, easily applied, but crude method while the second, based on the use of frequency-area curves, is more sophisticated but more difficult to apply, and involves a considerable amount of computation.

Frequency adjustment

7.2 The first method of adjustment arises from consideration of the main effect of altering quadrat size, which is to reduce or increase the frequency of occurrence of all species. Section 4 examined the detailed modelling of frequency-area curves for individual species and found that rates of change vary from species to species. However as a first approximation it may be sufficient to adjust the frequency of all species by the same factor.

7.3 Suppose it is required to compare two sets of quadrats obtained using different quadrat sizes. The average frequency per species will typically vary between the two sets and much of the difference may be due to the difference in quadrat size. One possible method of adjustment is to multiply the individual species frequencies obtained from the larger quadrats by the ratio of the average frequencies from the smaller and larger quadrats. This will reduce the average frequency of the larger quadrats to that of the set of smaller quadrats. The reason for scaling down the frequencies from the larger quadrats is that the alternative, scaling up the frequencies from the smaller quadrats, could produce values greater than one and this would complicate the ordination methodology.

7.4 Although this method seems initially plausible, it suffers from a major drawback. Larger quadrats will tend to pick up more species and, since these additional species are likely to be less common, they will tend to have lower frequencies. Thus the average frequency for the larger quadrats will be lower than needed to adjust for quadrat size. To overcome this the ratio of average frequencies could be calculated from only those species that occur in both sets of data. However if the two datasets represent very different vegetation types the set of common species may not be representative of either.

7.5 An alternative method is to use a scaling factor that is not so dependent on the number of species found. It is simple to show mathematically that the average frequency per species (f) for a set of quadrats is related to the average number of species per quadrat (n) and the total number of species observed (s) by the equation $f = n/s$. This equation makes clear the problem with using average frequency for observed species as a scaling factor. As quadrat size increases the number of species per quadrat will tend to increase but so will the number of species found overall. The average frequency will not therefore increase to the same extent since it is the ratio of two increasing numbers. It is even possible that average frequency might reduce with increasing quadrat size if the overall number of species increased more quickly than the number of species per plot. The alternative method is therefore to use the ratio of the average number of species per plot as the scaling factor in adjusting frequencies from the set of larger quadrats. This is equivalent to using the ratio of average frequencies when these are calculated using the combined set of species from both sets of quadrats, with species not found in one or other of the sets being given a frequency of zero for the set in which it does not occur.

7.6 Table 34 shows the average frequencies calculated in this way for the data from the Natural England sites. Several features are immediately apparent. First, as expected, the average frequency increases with nest size. This occurs for each site and for each NVC type. Secondly, also as expected, there is a gradual levelling off of frequency values at the larger nest sizes. This suggests that correction for quadrat size may not be necessary at larger quadrat sizes. However these are average values and it may be the case that the frequency of less common species is still increasing at large nest sizes. Thirdly there are significant differences between the average frequencies of the different NVC types at any specific nest size. The implication of this is that corrections made using the method suggested here will either over- or under-estimate the correct average frequency for the adjusted dataset to some degree. The extent to which this invalidates the method will depend on the size of the discrepancy in relation to the size of the adjustment.

Table 34 EN data - Average Frequency per species by site, NVC type and nest size

NVC type	Site		Nest									
			1	2	3	4	5	6	7	8	9	All
CG2	Bowerchalk Downs		0.049	0.071	0.087	0.106	0.123	0.135	0.152	0.164	0.184	0.119
	Knighton Bank		0.065	0.093	0.112	0.135	0.156	0.168	0.186	0.205	0.224	0.149
	Knighton Down		0.065	0.084	0.105	0.133	0.153	0.168	0.185	0.191	0.201	0.143
	Parsonage Down		0.050	0.067	0.091	0.119	0.141	0.158	0.174	0.189	0.203	0.132
	Pewsey Downs		0.056	0.072	0.088	0.107	0.127	0.147	0.168	0.180	0.193	0.126
	Wylye & Church Dean Down		0.059	0.085	0.105	0.132	0.156	0.175	0.198	0.210	0.225	0.150
	All sites	**mean**	0.057	0.079	0.098	0.122	0.143	0.158	0.177	0.190	0.205	0.137
		std dev.	0.007	0.010	0.011	0.013	0.015	0.015	0.016	0.017	0.017	0.013
CG5	Cotswolds Commons		0.048	0.065	0.076	0.097	0.112	0.121	0.135	0.145	0.159	0.107
	Edge Common		0.030	0.043	0.055	0.070	0.086	0.095	0.110	0.121	0.139	0.083
	Swifts Hill		0.033	0.045	0.058	0.073	0.088	0.097	0.112	0.124	0.137	0.085
	All sites	**mean**	0.037	0.051	0.063	0.080	0.095	0.104	0.119	0.130	0.145	0.092
		std dev.	0.010	0.012	0.011	0.015	0.014	0.015	0.014	0.013	0.012	0.013
MG3	Borrowbeck: Big Field		0.055	0.072	0.088	0.112	0.132	0.153	0.168	0.183	0.199	0.129
	Borrowbeck: By River		0.065	0.082	0.100	0.119	0.137	0.152	0.165	0.177	0.190	0.132
	Bowber Head and Piper Hole		0.046	0.063	0.079	0.096	0.108	0.124	0.138	0.152	0.167	0.108
	Muker Meadow: The Rash		0.041	0.057	0.070	0.085	0.095	0.109	0.123	0.137	0.157	0.097
	Muker Meadow: Yellands		0.054	0.068	0.080	0.091	0.102	0.114	0.121	0.131	0.137	0.100
	Pry and Bottom Meadow		0.063	0.077	0.091	0.109	0.122	0.136	0.153	0.164	0.176	0.121
	All sites	**mean**	0.054	0.070	0.085	0.102	0.116	0.131	0.145	0.157	0.171	0.115
		std dev.	0.009	0.009	0.011	0.013	0.017	0.019	0.021	0.021	0.022	0.015
	All	**mean**	0.052	0.070	0.086	0.106	0.123	0.137	0.152	0.165	0.179	0.119
		std dev.	0.011	0.014	0.016	0.020	0.024	0.026	0.028	0.029	0.029	0.021

7.7 This method of adjustment was applied to the two datasets used in Section 6 to illustrate the effects of mixing data from different nest sizes. The results are shown in Figures 48 and 49, which should be contrasted with Figures 33 and 34 respectively. Comparing these two pairs of figures with Figure 27 shows that in both examples the adjustment method has largely eliminated the distortion induced by the use of different nest sizes and has recovered the clear difference between the three NVC types.

7.8 Table 35 gives the average frequencies for the ADAS data. As with the Natural England dataset frequency increases steadily with nest size. There are also differences between the individual sites though these are relatively small except for the South Wessex Downs where the frequency is approximately twice that of the other sites. It is likely that this difference is the cause of this site appearing as an outlier in each of the individual nest plots of Figure 36. In effect the first correspondence analysis axis at each nest size is merely reflecting the difference in average frequency of one site from the others and does not reflect any difference in composition.

7.9 Figure 50 shows the result of applying the frequency adjustment method to the ADAS mixed nest size example presented in Figure 42. Comparison of these two figures with Figure 36 shows that, as with the Natural England examples, the adjusted dataset gives results much closer to those obtained from the single nest sizes.

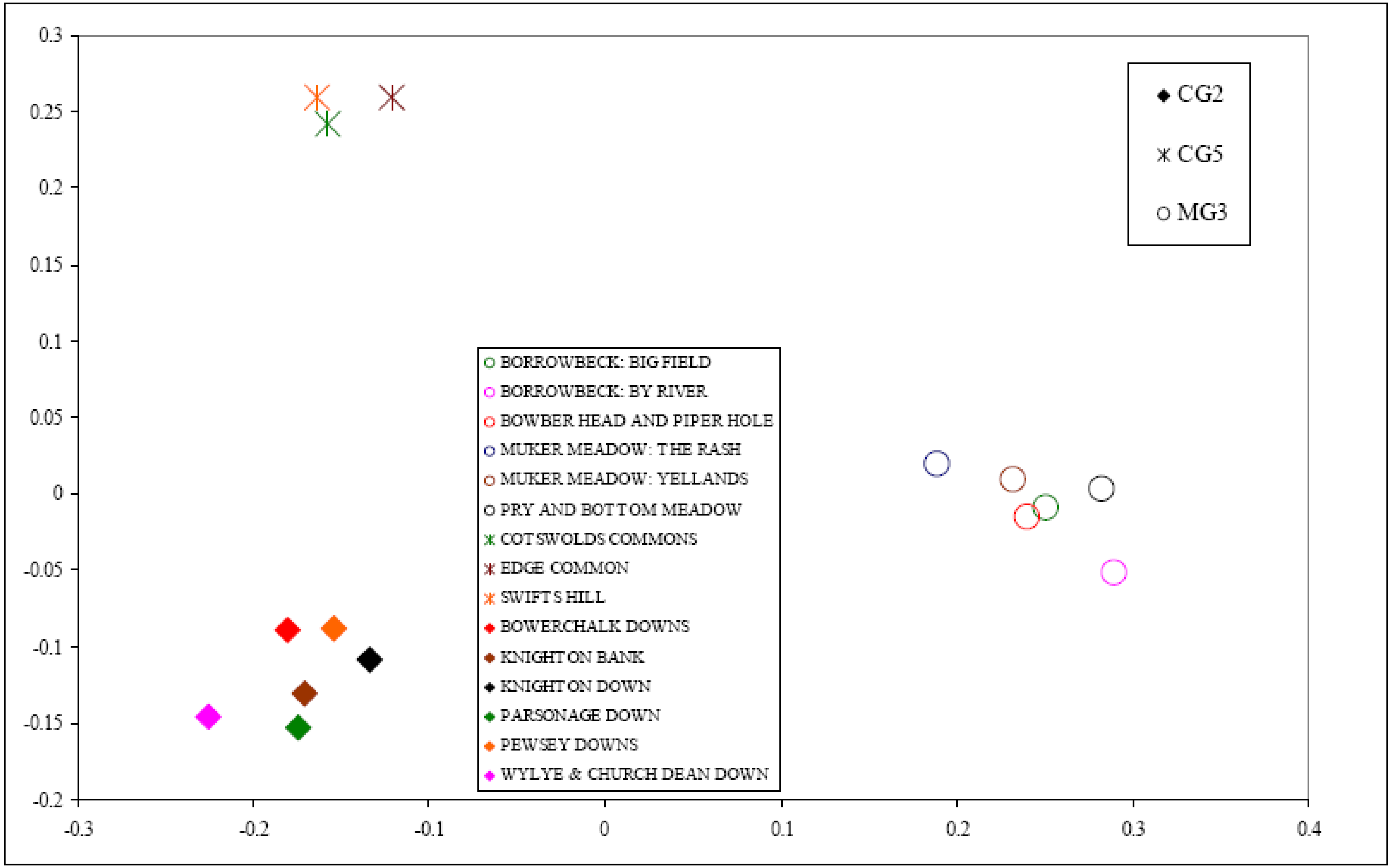


Figure 48 Natural England data Correspondence analysis dimensions 1 and 2 MG5, CG2 sites with smallest nests, CG5 sites with largest nests adjusted using average frequency in each nest size

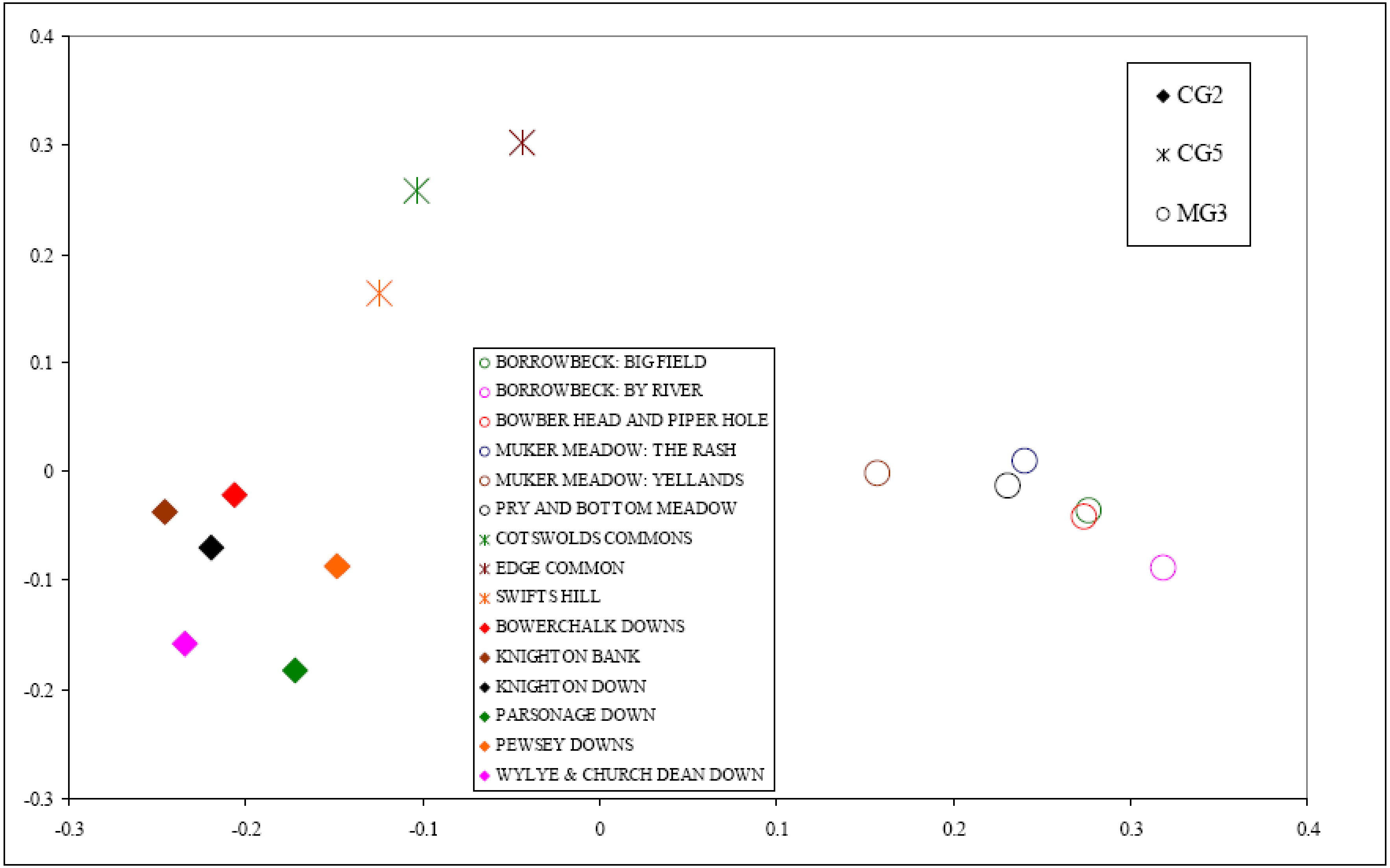


Figure 49 Natural England data Correspondence analysis dimensions 1 and 2 nest size varied across NVC category - Adjusted using average frequencies in each nest size

Table 35 ADAS ESA data - Average Frequency per species by site, activity and nest size

Site		Nest																
		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	All
Avon Valley	1	0.007	0.011	0.013	0.015	0.017	0.019	0.021	0.023	0.027	0.036	0.042	0.047	0.050	0.052	0.054	0.055	0.031
	2	0.007	0.009	0.011	0.013	0.015	0.018	0.021	0.022	0.024	0.035	0.043	0.050	0.053	0.055	0.057	0.059	0.031
Blackdown Hills		0.008	0.011	0.013	0.016	0.019	0.022	0.025	0.028	0.032	0.042	0.050	0.055	0.059	0.061	0.062	0.064	0.036
Clun ESA	1	0.007	0.009	0.011	0.012	0.015	0.017	0.019	0.021	0.023	0.029	0.033	0.038	0.040	0.042	0.043	0.044	0.025
	2	0.010	0.012	0.015	0.016	0.019	0.021	0.023	0.025	0.028	0.037	0.042	0.046	0.048	0.049	0.051	0.052	0.031
Dartmoor	1	0.008	0.010	0.011	0.013	0.015	0.016	0.018	0.020	0.022	0.028	0.032	0.034	0.036	0.037	0.037	0.038	0.023
	2	0.013	0.017	0.020	0.025	0.027	0.030	0.035	0.039	0.042	0.050	0.055	0.059	0.060	0.062	0.063	0.063	0.041
Exmoor	1	0.009	0.011	0.012	0.014	0.015	0.017	0.019	0.021	0.023	0.031	0.036	0.040	0.041	0.042	0.044	0.045	0.026
	2	0.010	0.012	0.015	0.018	0.021	0.025	0.028	0.030	0.034	0.046	0.054	0.060	0.064	0.067	0.069	0.071	0.039
Lake District		0.010	0.012	0.015	0.017	0.021	0.024	0.027	0.030	0.034	0.048	0.055	0.060	0.062	0.064	0.065	0.066	0.038
North Peak		0.007	0.009	0.011	0.013	0.015	0.018	0.020	0.023	0.025	0.031	0.036	0.040	0.042	0.044	0.044	0.045	0.027
Shropshire Hills		0.009	0.012	0.014	0.017	0.019	0.022	0.024	0.027	0.030	0.038	0.044	0.048	0.050	0.052	0.053	0.055	0.032
Somerset Levels and Moors		0.007	0.009	0.012	0.015	0.018	0.021	0.023	0.027	0.031	0.040	0.047	0.052	0.054	0.056	0.057	0.058	0.033
South Wessex Downs		0.016	0.023	0.028	0.034	0.040	0.047	0.053	0.059	0.065	0.082	0.094	0.101	0.104	0.106	0.108	0.110	0.067
South West Peak	1	0.009	0.012	0.014	0.016	0.017	0.019	0.022	0.024	0.026	0.034	0.039	0.043	0.045	0.047	0.049	0.051	0.029
	2	0.010	0.013	0.016	0.019	0.022	0.024	0.027	0.029	0.032	0.041	0.047	0.051	0.053	0.054	0.055	0.056	0.034
Upper Thames Tributaries		0.010	0.013	0.016	0.018	0.021	0.023	0.025	0.028	0.030	0.039	0.046	0.050	0.053	0.054	0.056	0.056	0.034
West Penwith		0.005	0.006	0.008	0.010	0.012	0.014	0.016	0.019	0.022	0.030	0.035	0.039	0.041	0.042	0.043	0.044	0.024
All	**mean**	0.009	0.012	0.014	0.017	0.019	0.022	0.025	0.027	0.030	0.040	0.046	0.051	0.053	0.055	0.056	0.057	0.033
	std dev.	0.002	0.003	0.004	0.005	0.006	0.007	0.008	0.009	0.010	0.012	0.014	0.015	0.015	0.015	0.016	0.016	0.010

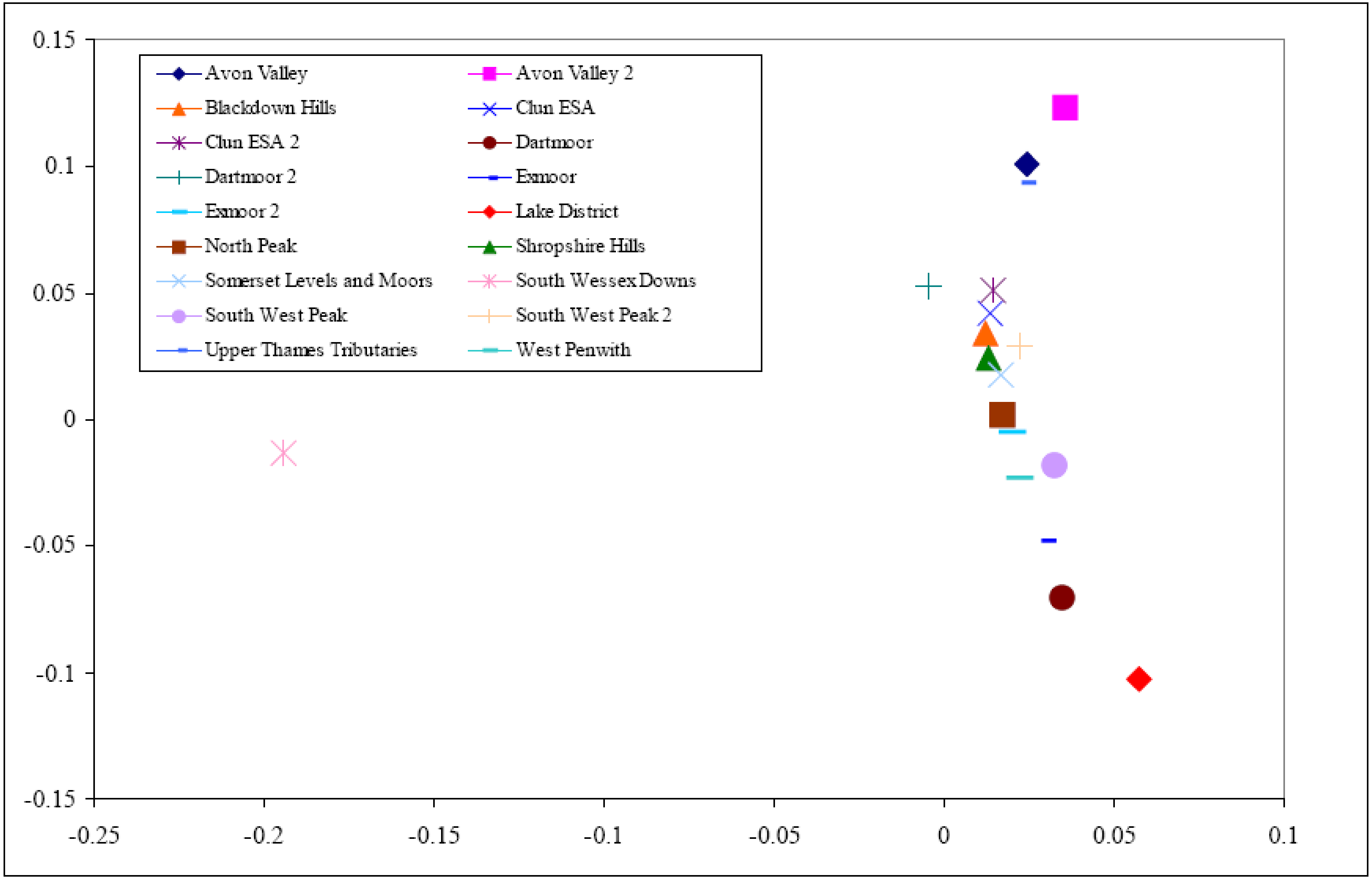


Figure 50 ADAS data Correspondence analysis dimensions 1 and 2 nest size varied with site - Adjusted using average frequencies in each nest size

Table 36 Countryside Survey data - average frequency per species by aggregate class and nest size

CVS Aggregate class		Nest					
		1	2	3	4	5	All
AG0		0.002	0.002	0.003	0.003	0.003	0.003
AG1		0.006	0.007	0.007	0.008	0.008	0.007
AG2		0.005	0.007	0.008	0.009	0.010	0.008
AG3		0.009	0.011	0.012	0.014	0.015	0.012
AG4		0.016	0.020	0.023	0.025	0.027	0.022
AG5		0.008	0.012	0.015	0.017	0.020	0.014
AG6		0.008	0.011	0.013	0.015	0.018	0.013
AG7		0.015	0.020	0.023	0.026	0.029	0.023
AG8		0.011	0.013	0.015	0.017	0.019	0.015
All	**mean**	0.009	0.011	0.013	0.015	0.017	0.013
	std. dev.	0.004	0.006	0.006	0.007	0.008	0.006

7.10 Table 36 gives the average frequencies for the CS dataset. While frequency increases with nest size the change is less than for the other two datasets, reflecting the fact that the CS nest sizes start at a larger size. Differences between aggregate classes are much greater than the differences between sites found in the other two datasets.

7.11 Figure 51 shows the result of applying the frequency adjustment method to the CS mixed nest size example presented in Figure 46. Unlike the examples from the other two datasets the CS adjusted dataset is not closer to the ordinations of the individual nests (Figure 43). Thus the method of adjustment appears not to have worked in this instance. The reason is the disparity in average frequency between the aggregate classes (Table 35). The CS mixed nest size example comprises AG0, AG1, AG2 and AG5 from nest 1 and the remaining aggregate classes from nest 5. The average frequency in the first group is approximately twice that of the second group, overall and in each individual nest. In reducing the frequency of the second group to that of the first the adjustment method has over-corrected for the disparity in nest size. This has a number of consequences, illustrated in Figure 52. Figure 52 contains three plots. The top plot is the adjusted dataset (with the horizontal axis reversed for comparability) and the bottom plot is the data from nest 1. The centre plot is the example dataset adjusted in a different way. To produce this plot the nest 5 classes have had their average frequency adjusted to the average frequency of the same classes in nest 1, and not to the average frequency of the nest 1 classes in the example dataset. Thus this adjustment has not over-corrected for nest size. The centre and bottom plots are very similar showing that the adjustment method is effective when it does not over-correct. The top plot is not too dissimilar to the other two at first sight, however the bottom two plots are of dimensions 1 and 3, not dimensions 1 and 2 as in the top plot. Thus the first consequence of the over-correction is an alteration in the order of the ordination axes. In addition the over-correction has produced a relative inflation of the frequencies in classes AG0, AG1 and AG2 and this results in their shift from the centre of axis 1 (in the bottom plot) to the left hand side (in the top plot).

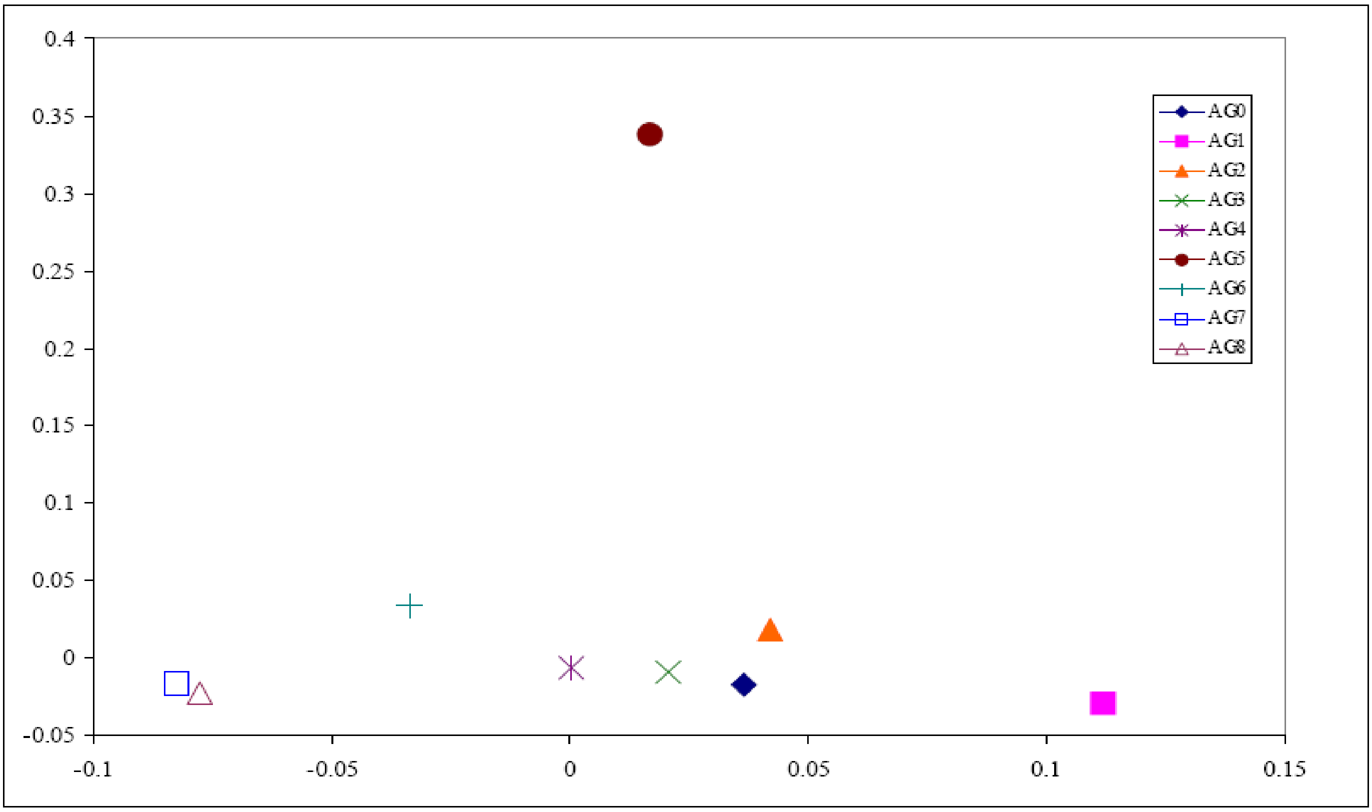


Figure 51 CS data Correspondence analysis dimensions 1 and 2 AG0, AG1, AG2 AG5 nest size 1: AG3, AG4, AG6, AG7, AG8 nest size 5 Adjusted using average frequencies in each nest size

7.12 Overall therefore the extent to which the proposed adjustment method is successful in correcting for differences in quadrat size depends on the accuracy with which it scales average frequencies. As applied here frequencies are scaled using the average frequency from all sites or classes within each nest size group. This retains the relative differences between sites within each size group but eliminates overall differences between size groups. Whether this is appropriate will in general be unknown An alternative is to scale each site or class separately thus completely eliminating differences in average frequency between sites and groups of sites. Results from such an analysis would presumable reflect just differences in composition and not differences in average frequency.

7.13 Figures 53 to 56 examine the difference between these two methods of adjustment for the Natural England dataset. Figures 53 and 54 show the first four dimensions of the correspondence analysis obtained by adjusting sites using the average frequency in each nest and Figures 55 and 55 show the results of adjusting each site separately. There is little difference in the first two dimensions. For both methods and for all nests the data separates into three groups of points representing the division of sites into NVC types. Hence the difference between types is clearly a difference in composition and not frequency. Dimensions 3 and 4 show more differences. Adjusting by site reverses the order of these two dimensions compared to adjusting by nest and the placing of some sites has changed substantially. In particular the site, Muker Meadow (The Rash) is much further from the other sites.

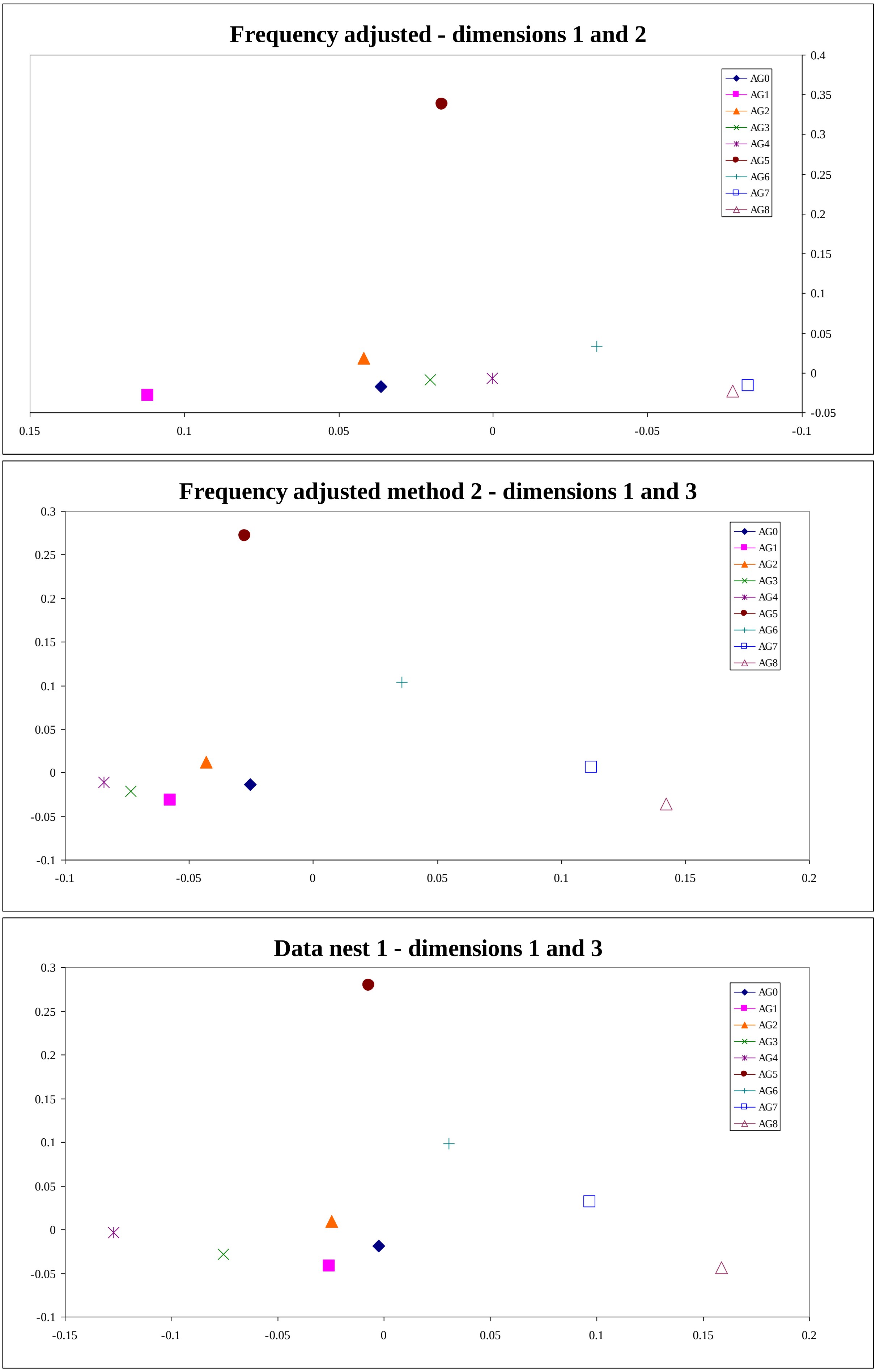


Figure 52 CS data Correspondence analysis Mixed nest size example

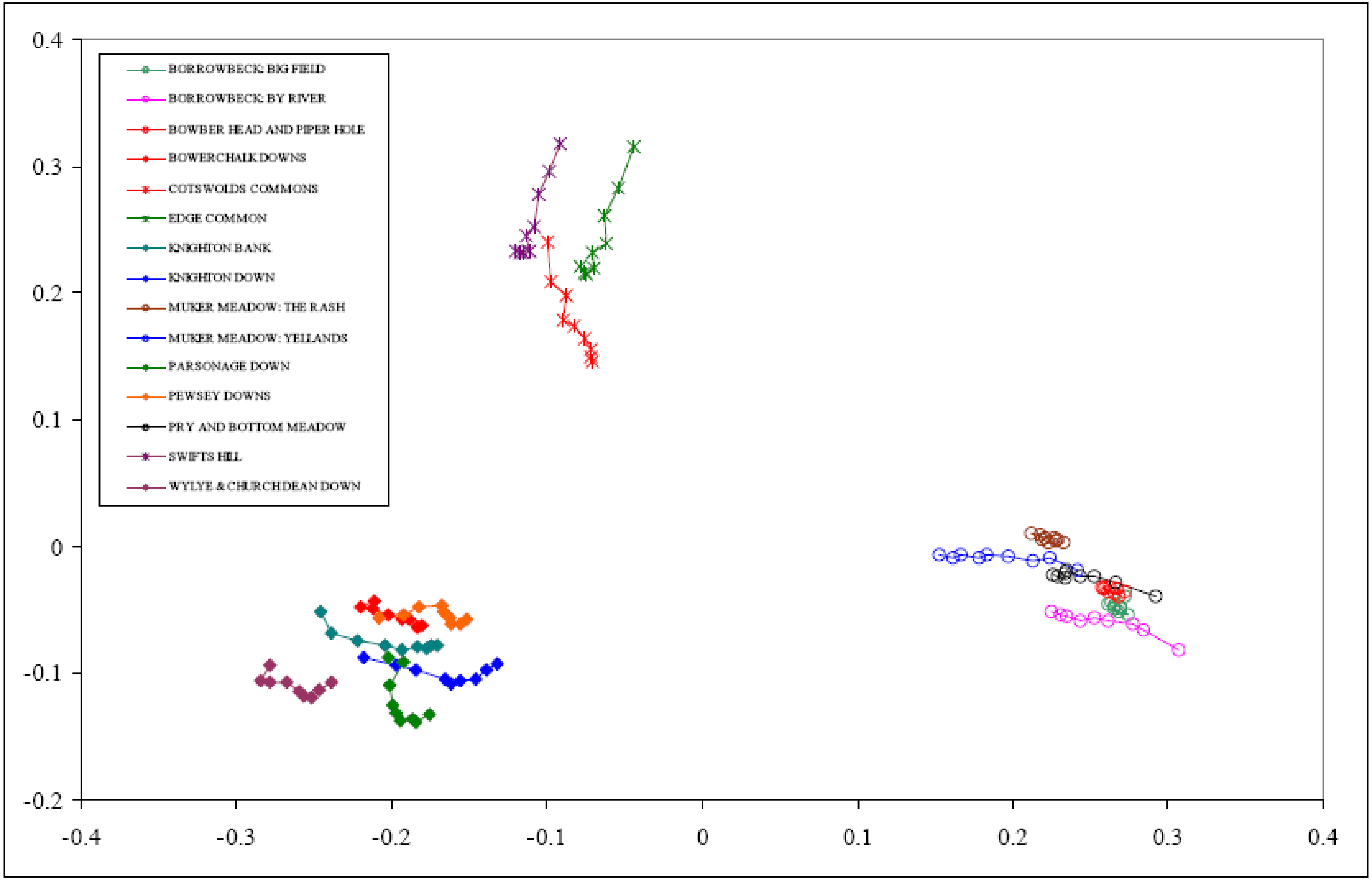


Figure 53 Natural England data Correspondence analysis dimensions 1 and 2, nests entered as separate points - adjusted using average frequency in each nest

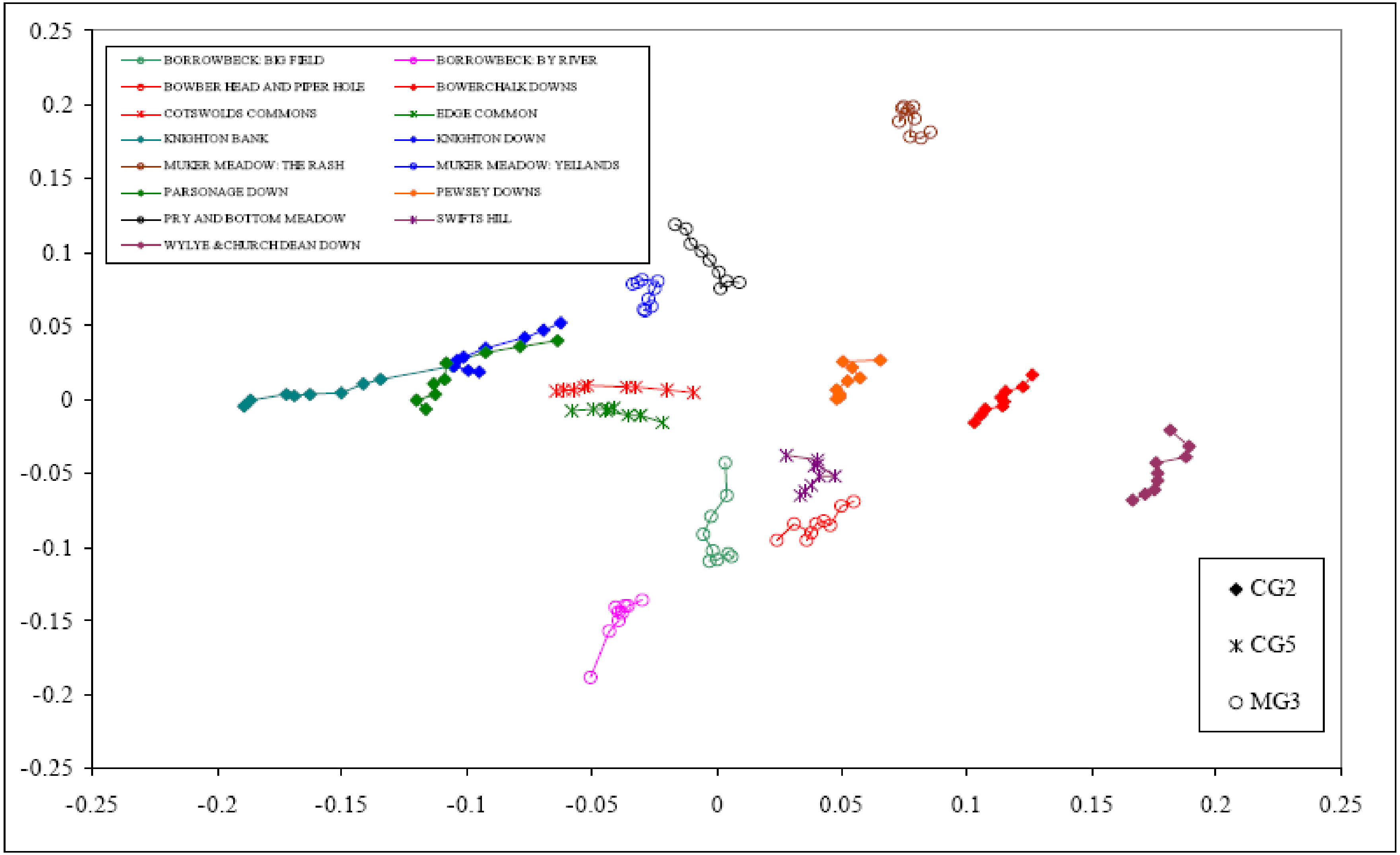


Figure 54 Natural England data Correspondence analysis dimensions 3 and 4, nests entered as separate points - adjusted using average frequency in each nest

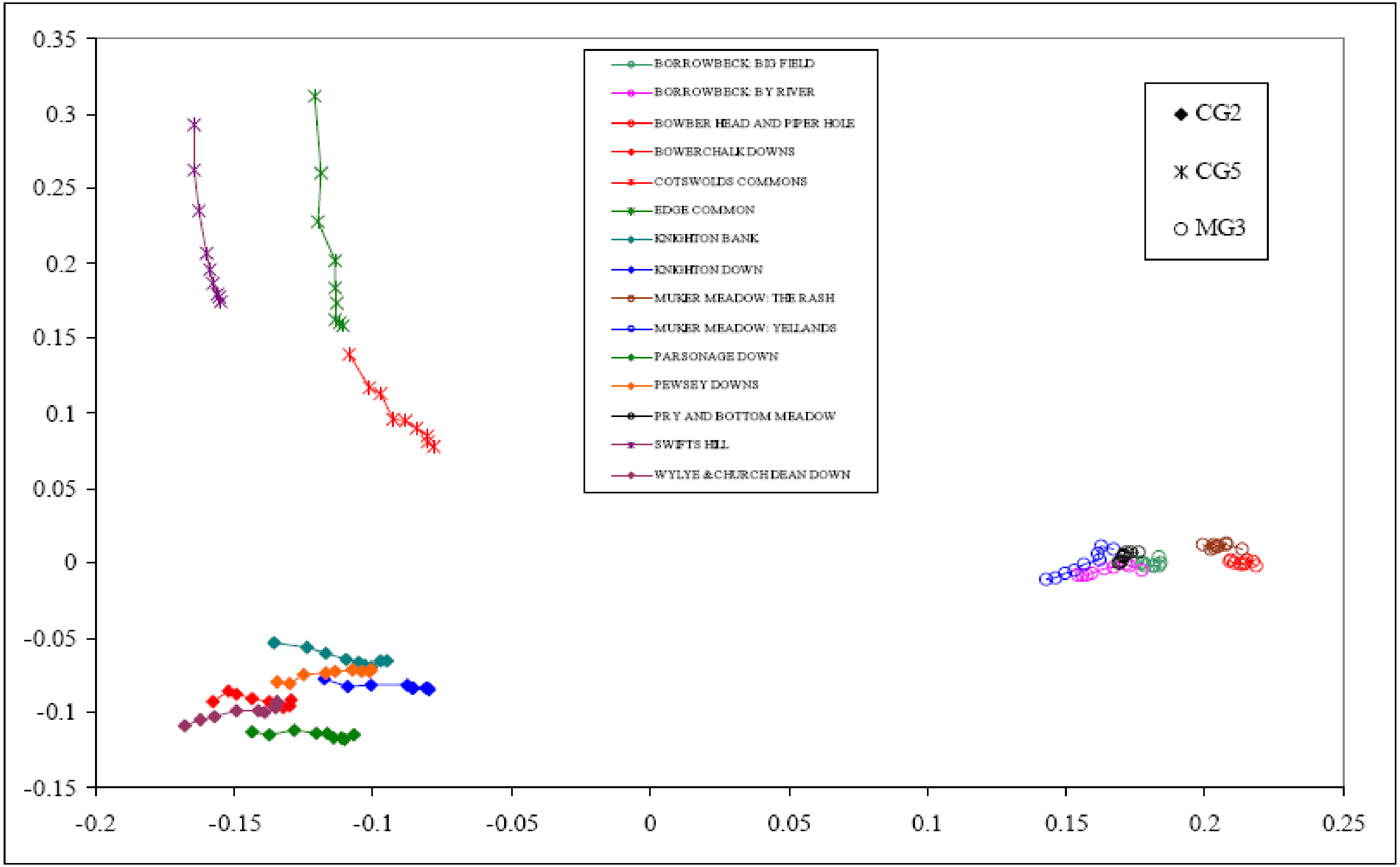


Figure 55 Natural England data Correspondence analysis dimensions 1 and 2, nests entered as separate points - adjusted using average frequency at each site

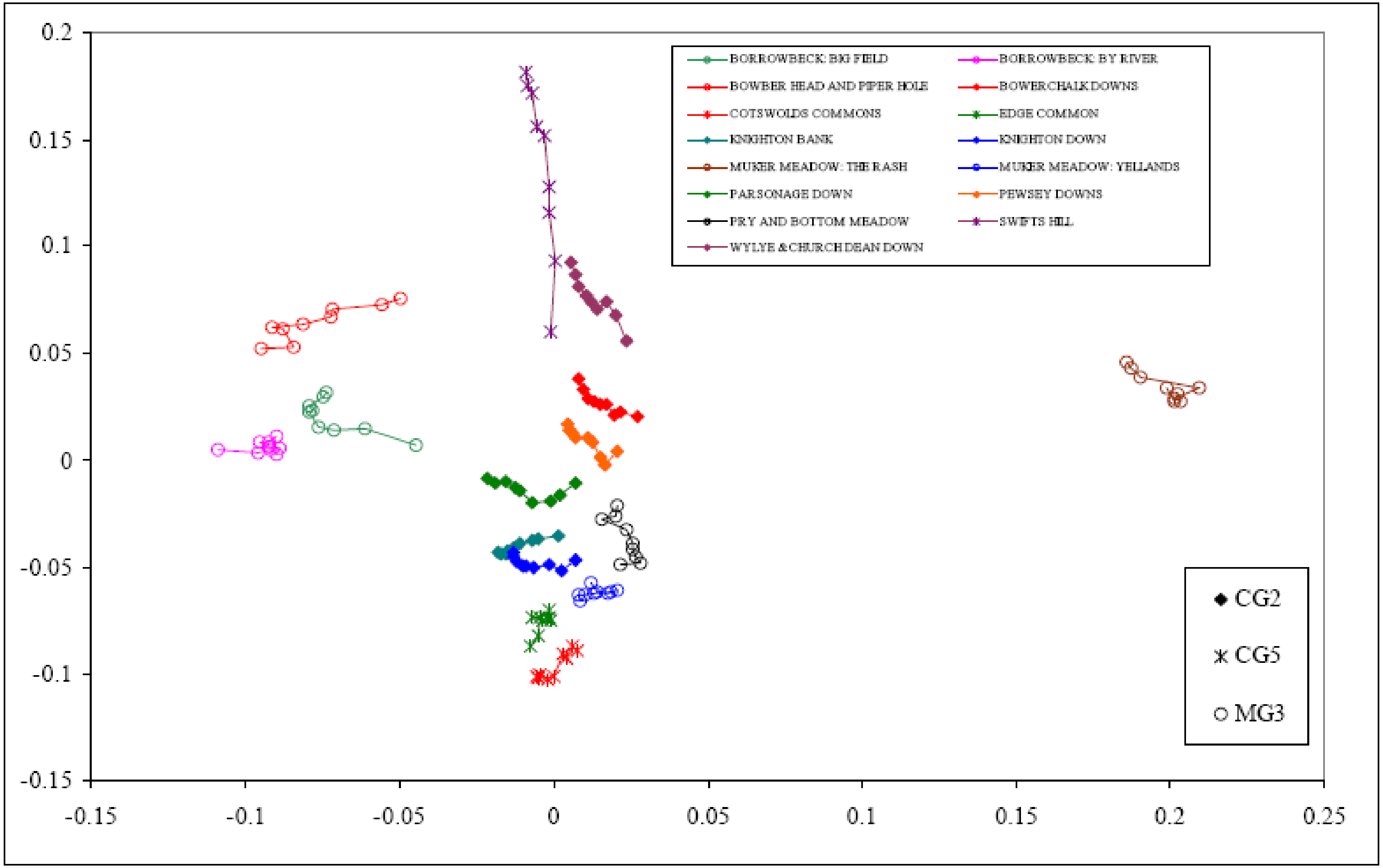


Figure 56 Natural England data Correspondence analysis dimensions 3 and 4, nests entered as separate points - adjusted using average frequency at each site

7.14 Figures 57 to 60 show the two methods of adjustment applied to the ADAS data. In this case it is the first two dimensions that show differences between the two adjustment methods (Figures 57 and 59) while dimensions 3 and 4 are essentially the same (Figures 58 and 60). When adjusting by average nest frequency (Figure 57) the first axis is dominated by the difference between the South Wessex Downs and the other sites while a major factor on axis 2 is that the Lake District is separated from the other sites. In contrast when adjusting by individual site frequencies (Figure 59) the difference between the South Wessex Downs and the other sites is reduced to such an extent that it forms the second axis while the Lake District is no longer separated from the other sites. Table 35 shows the reason for the change in the positioning of the South Wessex Downs. This site has an average frequency approximately twice that of all the other sites so that its extreme position in Figure 57 is at least partly due to its greater number of species per quadrat rather than differences in composition. The Lake District on the other hand has a similar average frequency to other sites so that its separation from the other sites must be due solely to differences in composition. It is separated from the other sites using both adjustment methods but on different axes. When adjustment is made using individual site frequencies the separation is confined to axis 4 (Figure 60) while adjusting using nest frequency shows the separation on several axes.

7.15 Figures 61 to 64 show the results of applying the two methods of adjustment to the CS dataset. As pointed out above the differences in average frequency between the CS Aggregate Classes (Table 36) are much greater than between the sites in the other two datasets. It is not surprising therefore that the two methods of adjustment produce substantially different results. In particular Aggregate Classes 0, 1 and 2 have lower frequencies than the others. This difference is retained when frequencies are adjusted by nest resulting in these three classes being relatively close together in Figures 61 and 62. When adjustment is by site the differences in composition of these classes becomes clearer (Figures 63 and 64). Another feature of this dataset not occurring in the other two is that differences between nests at the same site are smaller when adjustment is by site instead of by nest. Thus each nest is providing exactly the same information about composition.

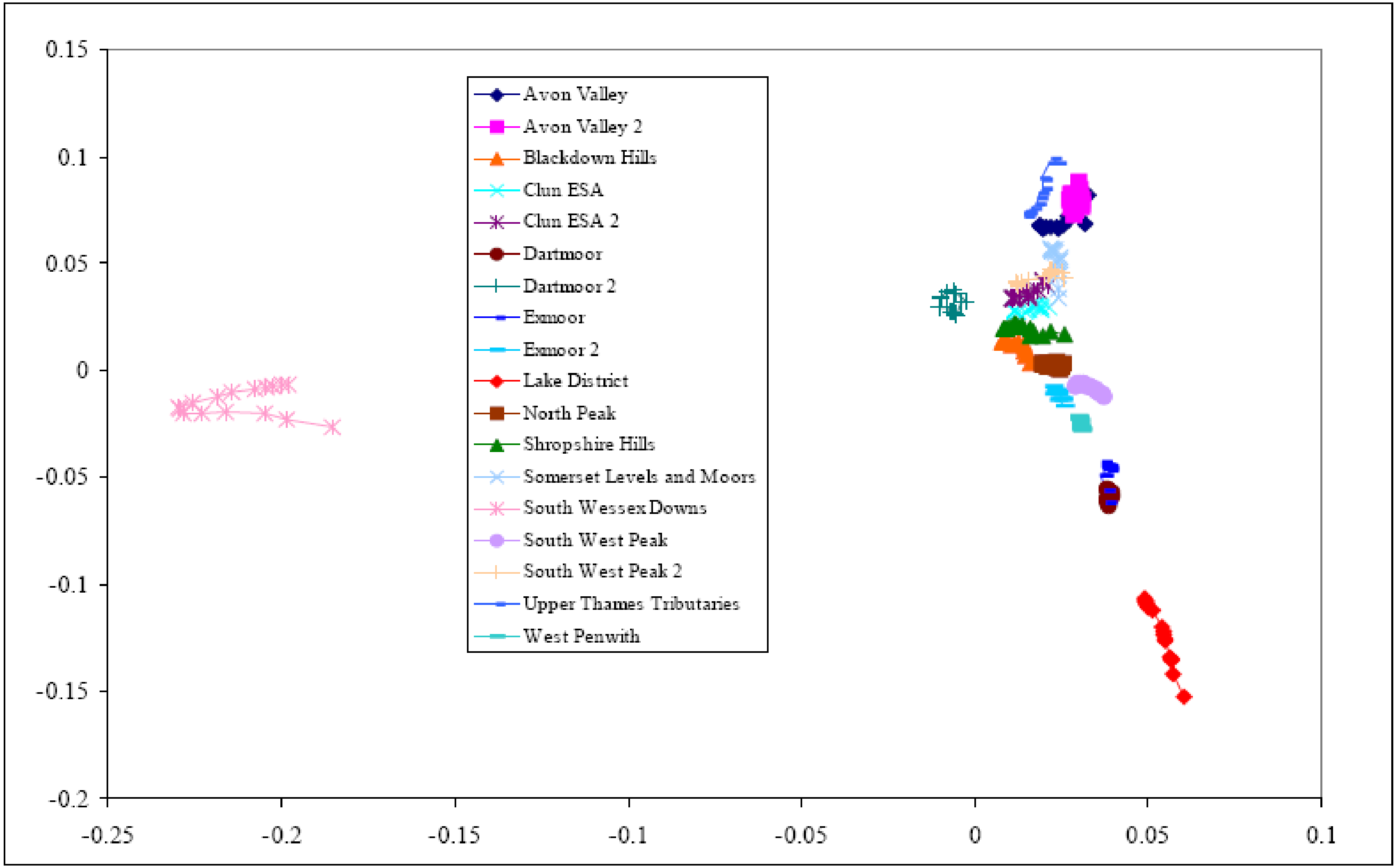


Figure 57 ADAS data Correspondence analysis dimensions 1 and 2, nests entered as separate points - adjusted using average frequency in each nest size

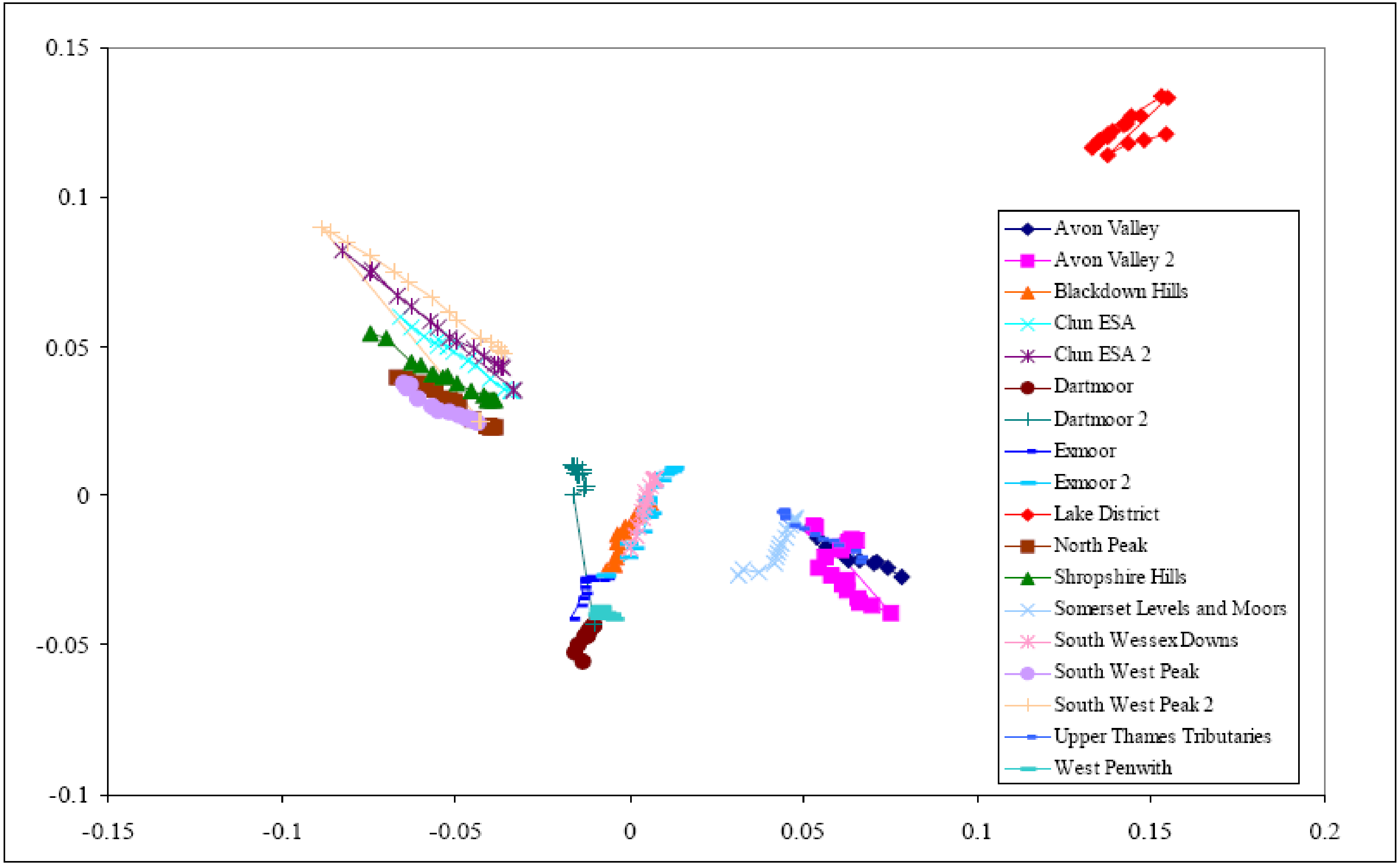


Figure 58 ADAS data Correspondence analysis dimensions 3 and 4, nests entered as separate points - adjusted using average frequency in each nest size

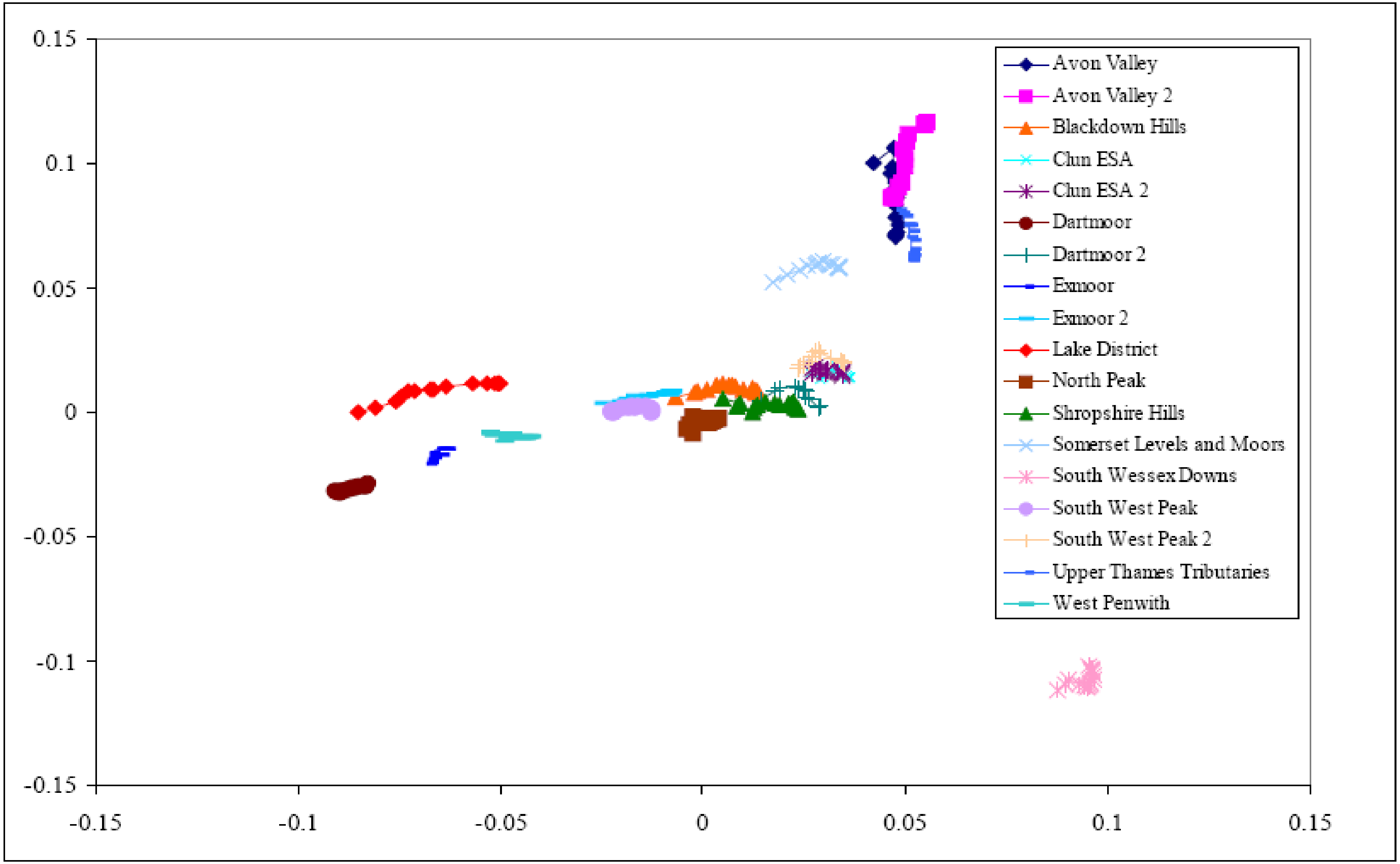


Figure 59 ADAS data Correspondence analysis dimensions 1 and 2, nests entered as separate points - adjusted using average frequency in each site

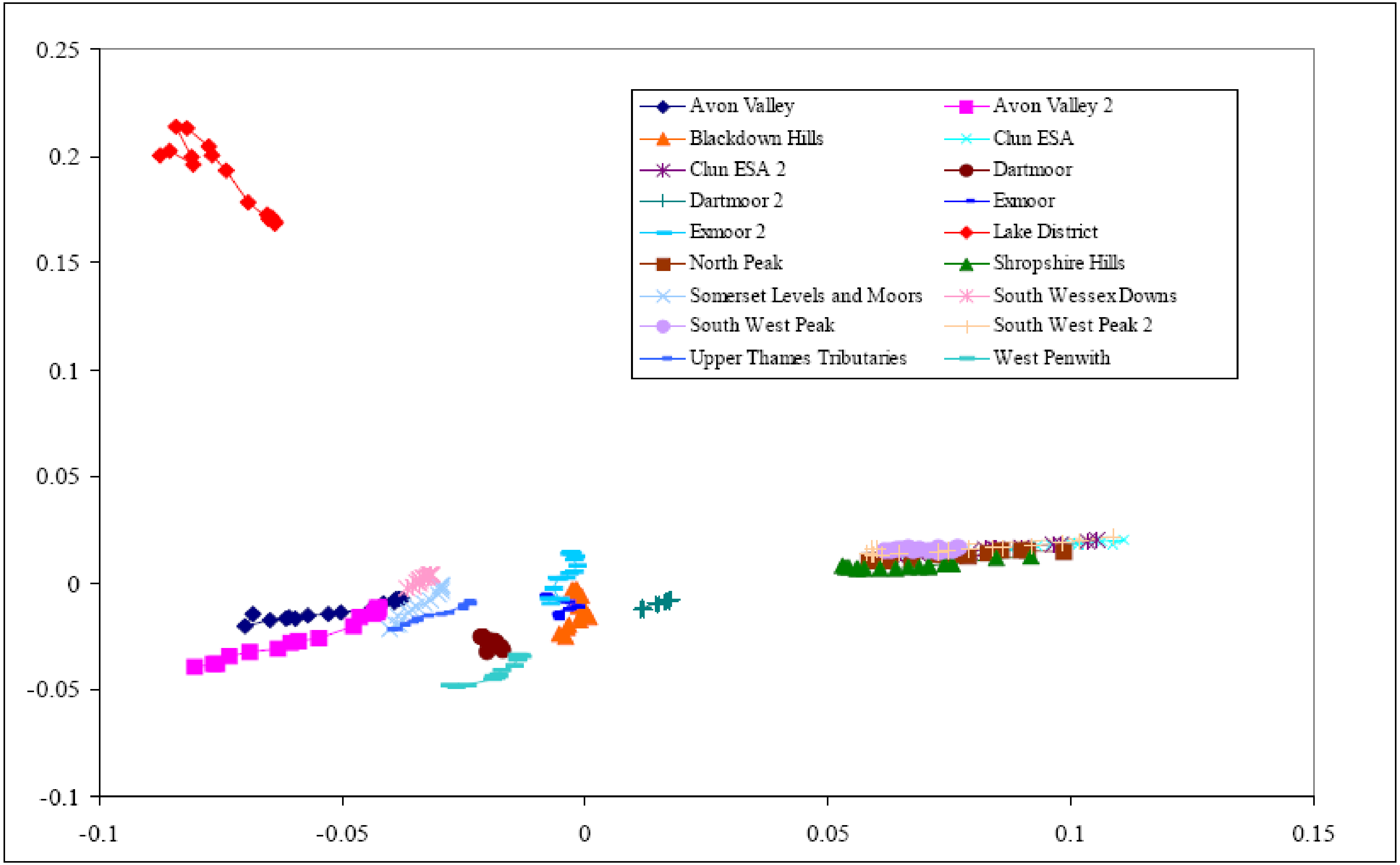


Figure 60 ADAS data Correspondence analysis dimensions 3 and 4, nests entered as separate points - adjusted using average frequency in each site

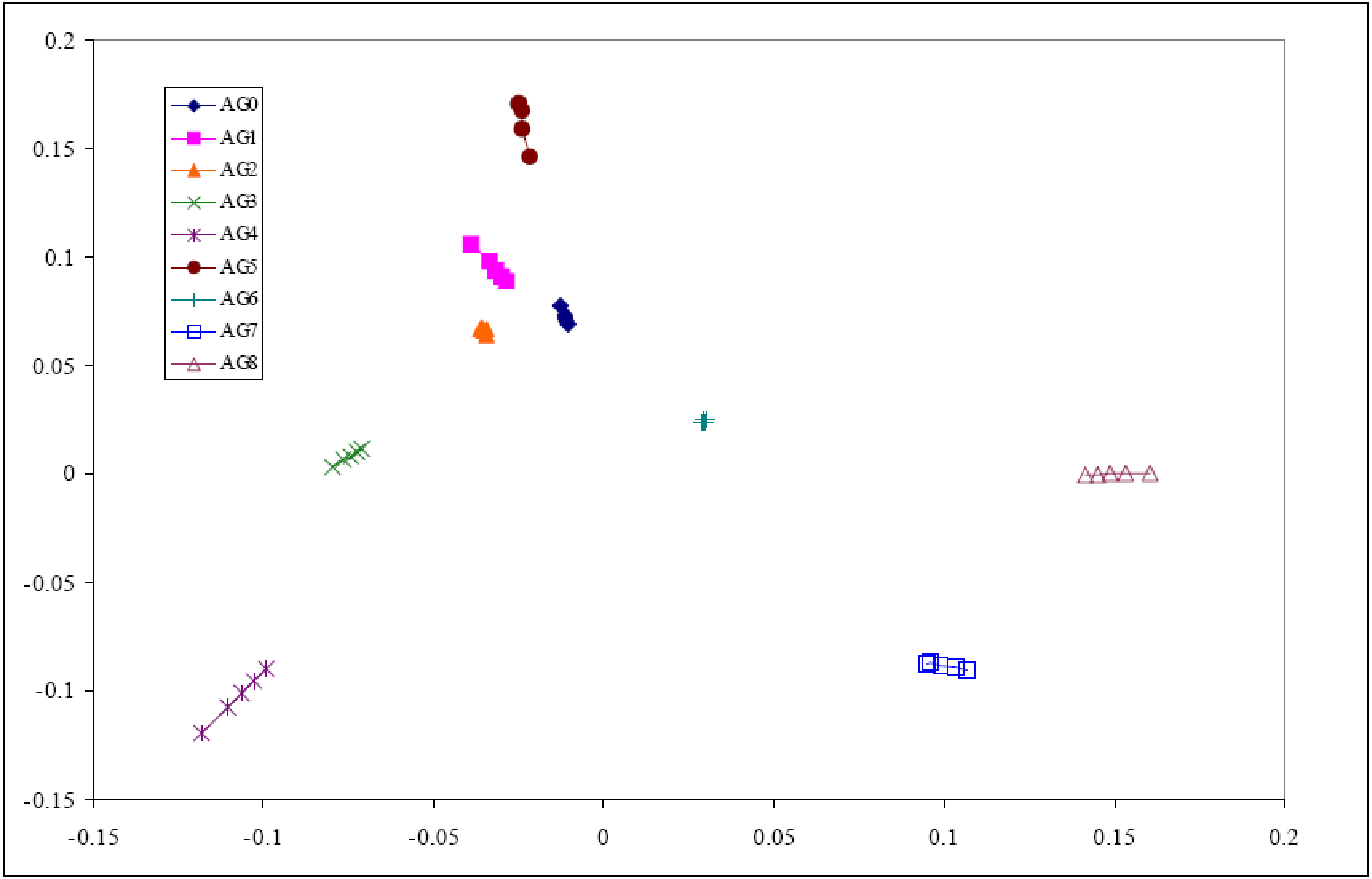


Figure 61 CS data Correspondence analysis dimensions 1 and 2, nests entered as separate points - adjusted using average frequency in each nest

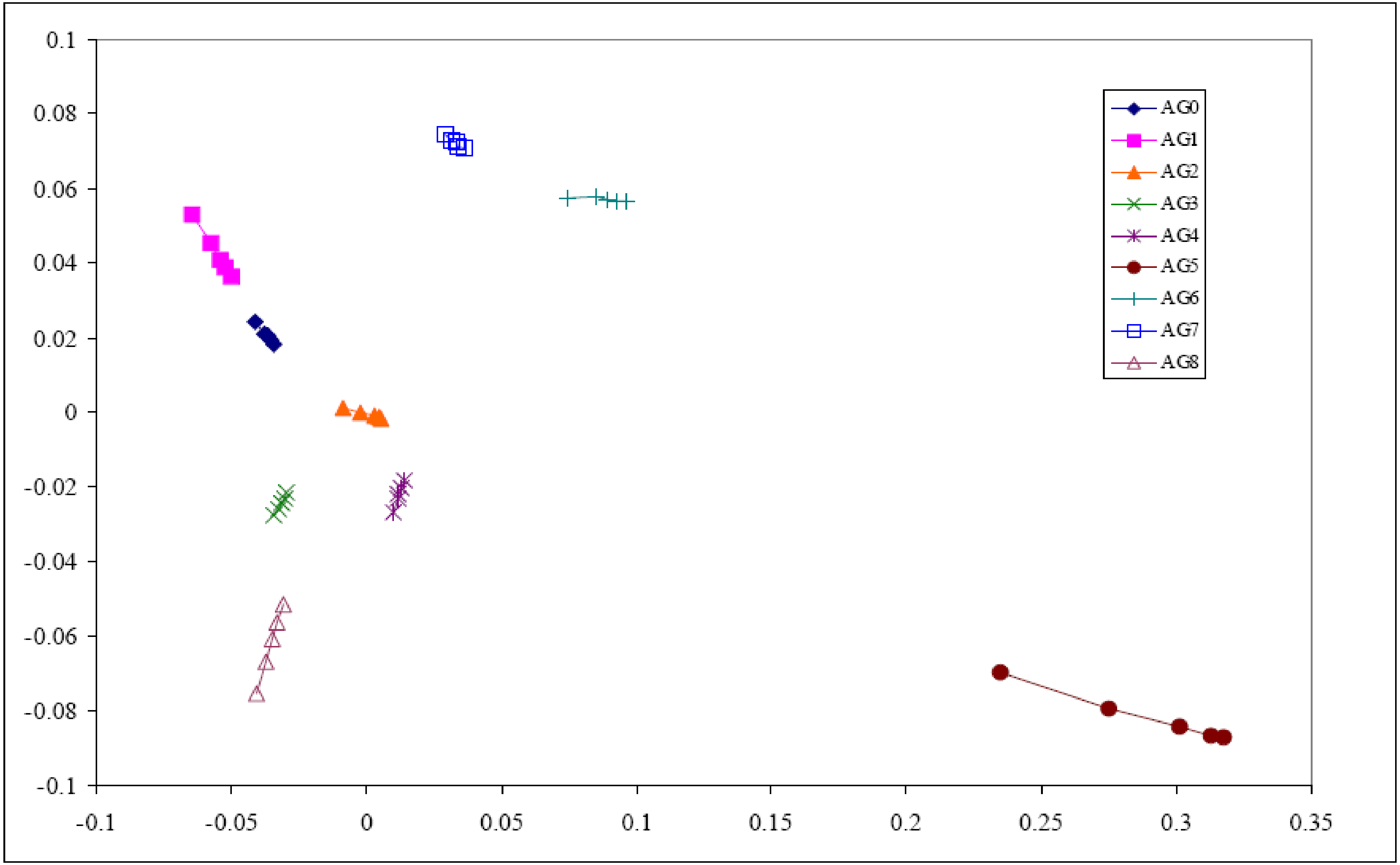


Figure 62 CS data Correspondence analysis dimensions 3 and 4, nests entered as separate points - adjusted using average frequency in each nest

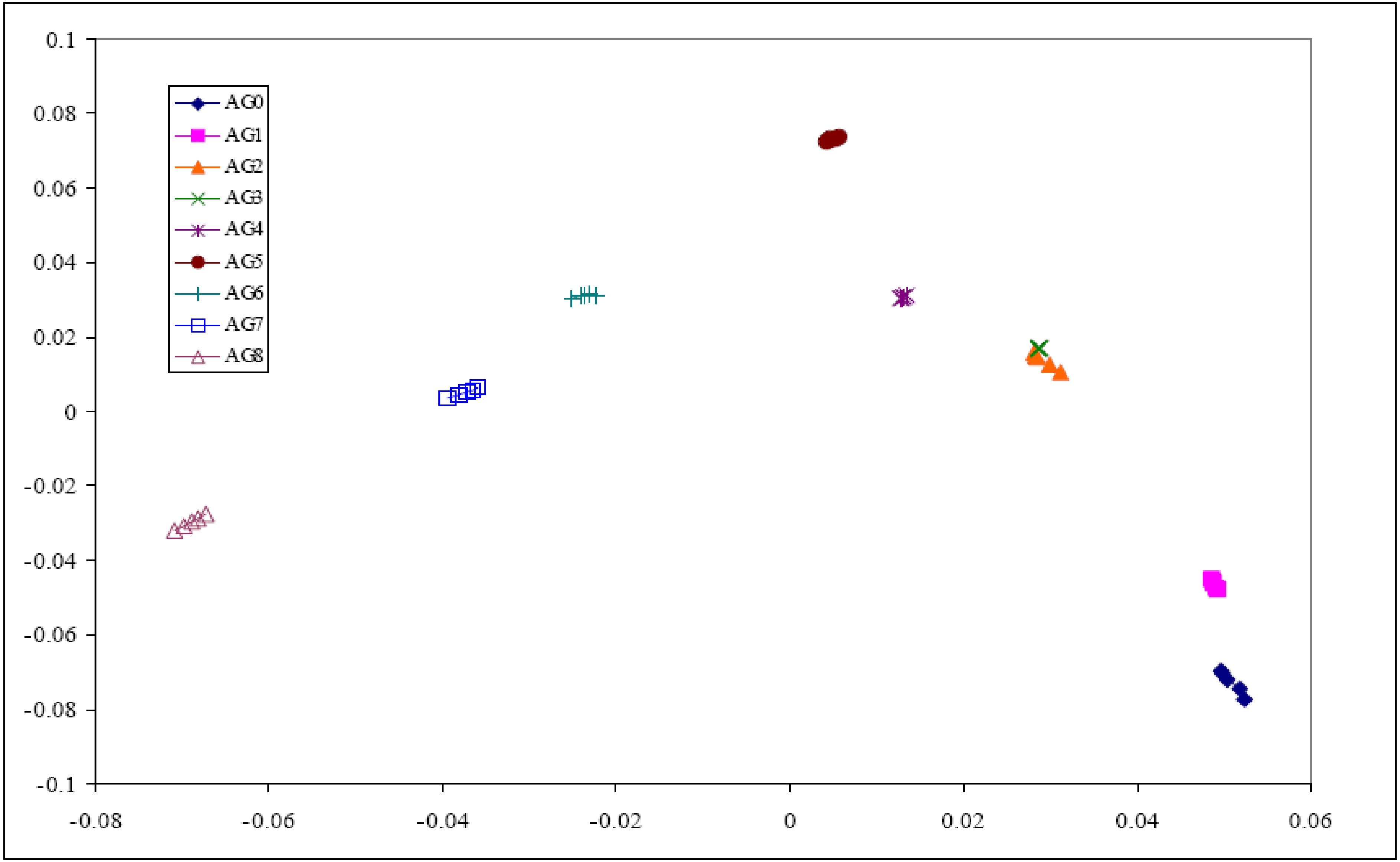


Figure 63 CS data Correspondence analysis dimensions 1 and 2, nests entered as separate points - adjusted using average frequency in each site

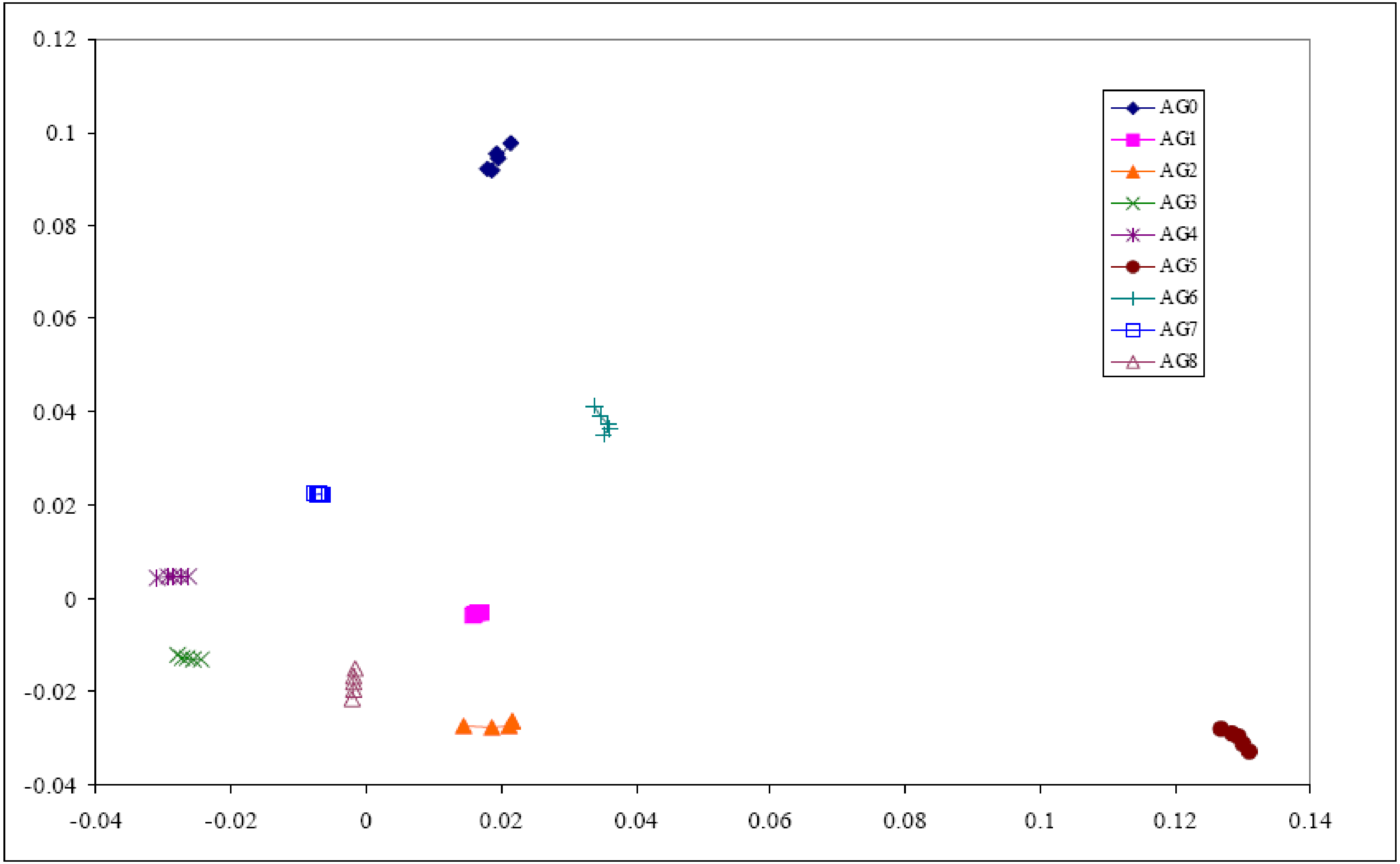


Figure 64 CS data Correspondence analysis dimensions 3 and 4, nests entered as separate points - adjusted using average frequency in each site

7.16 The methods of adjustment proposed above have a number of advantages. They are simple to apply and can be used with any dataset that can be reduced to presence/absence form. Since they do not utilise information on frequency-area relationships, they do not require nested data for implementation. It would seem therefore that the methods should have considerable utility in standardising datasets, particularly historic datasets where nested quadrats were not commonly used. The main drawback is that the extent to which the methods accurately correct for differences in quadrat size will generally be unknown. If, for example, one of the main differences between two vegetation types is the number of species per plot and the two types have been recorded using quadrats of different size then adjustment will eliminate this disparity and the resulting analytical comparison may, as a consequence, miss real differences. If, however, differences in composition are considered to be of primary importance then this method of adjustment is ideal and allows comparisons to be made regardless of quadrat size. Further work looking at variation in species per plot with vegetation type and other variables would throw light on this problem and might allow the method to be refined if necessary.

Using frequency-area curves

7.17 The second method of adjusting for differences in quadrat size between two datasets is to use fitted frequency-area curves to extrapolate for one or both of the datasets to a common quadrat size. Thus the method requires that at least one of the datasets must be based on the use of nested quadrats.

7.18 To illustrate this method we concentrate on the ADAS dataset since it has the widest range of nest sizes. The two subsets of this data formed from nests above 1 m^2 and from nests 1 m^2 and below were each used to predict frequencies at the omitted nest sizes. Since these two subsets approximate in range of nest sizes to the EN and CS datasets their results should reflect the effects of standardisation using these datasets.

7.19 Some difficulties arise in applying this technique. Firstly the question of what to do with species for which an adequate fit is not obtained, either because they are observed in an insufficient number of plots or because they are found in too many small nests. In Section 4 it was suggested that only for species found in twenty or more quadrats and six or more nests could an accurate choice of curve be made. If such an extreme criteria were adopted then a large proportion of species would not qualify. However in some cases difficulty in choosing a parameterisation arose because different curve parameterisations fitted equally well. For prediction, therefore, a less stringent condition might be appropriate.

7.20 Two different situations arise. If prediction is to be made for quadrat sizes above existing nest sizes then the difficulty is in making predictions for rare species which would be expected to rise in frequency. Choice of prediction method for species already found in virtually all plots is unlikely to have a major effect on analysis since such species will be ubiquitous at larger nest sizes and differences in predictions small. If prediction is to be made for quadrat sizes smaller than those already in use then the opposite situation arises. Prediction for ubiquitous species expected to drop in frequency is likely to be inaccurate whereas rare species will become rarer and less likely to influence community analysis to any extent.

7.21 The solution adopted here is a pragmatic one. Predictions are made for all species for which a curve can be fitted. For species where this is not the case the frequency at the smallest nest size is used when prediction is for a quadrat size below the existing nest sizes and the frequency at the largest nest size when prediction is for quadrat sizes above existing nest sizes.

7.22 Two possible alternatives are not examined in detail here. The first is to omit species for which curves can not be fitted. While this might produce sensible results in some circumstances there is a danger that important species may be omitted and a distorted analysis obtained. The second alternative is a hybrid approach using fitted frequency-area curves where these are considered adequate and the adjusted frequency approach, as described above, for the remaining species but basing the change in frequency on those species to which frequency-area curves have been fitted instead of differences in frequency of two datasets. In practise analyses can be obtained for a range of strategies. If no substantial differences arise then more confidence can be invested in the results.

7.23 A choice also has to be made of which form of frequency-area curve to use for each species. One option is to use the same form for all species. The advantage of this approach is that extrapolation outside the range of the observed data will produce more consistent predictions for the set of species as a whole. Alternatively the best fitting curve type could be used for each species but this is likely to produce inconsistent predictions, especially at the limits of extrapolation. In Section 4 it was shown that the *cd* and *acd* parameterisations of the logistic curve were the best choice of model, providing a good fit for almost all species. Attention is therefore restricted to these curves. The analyses presented here compare three methods of using these curves. First using the *cd* parameterisation for all species, second using the *acd* parameterisation for all species, and finally a mixed approach using the *cd* parameterisation for the majority of species and the *acd* parameterisation for those species for which it is significantly better.

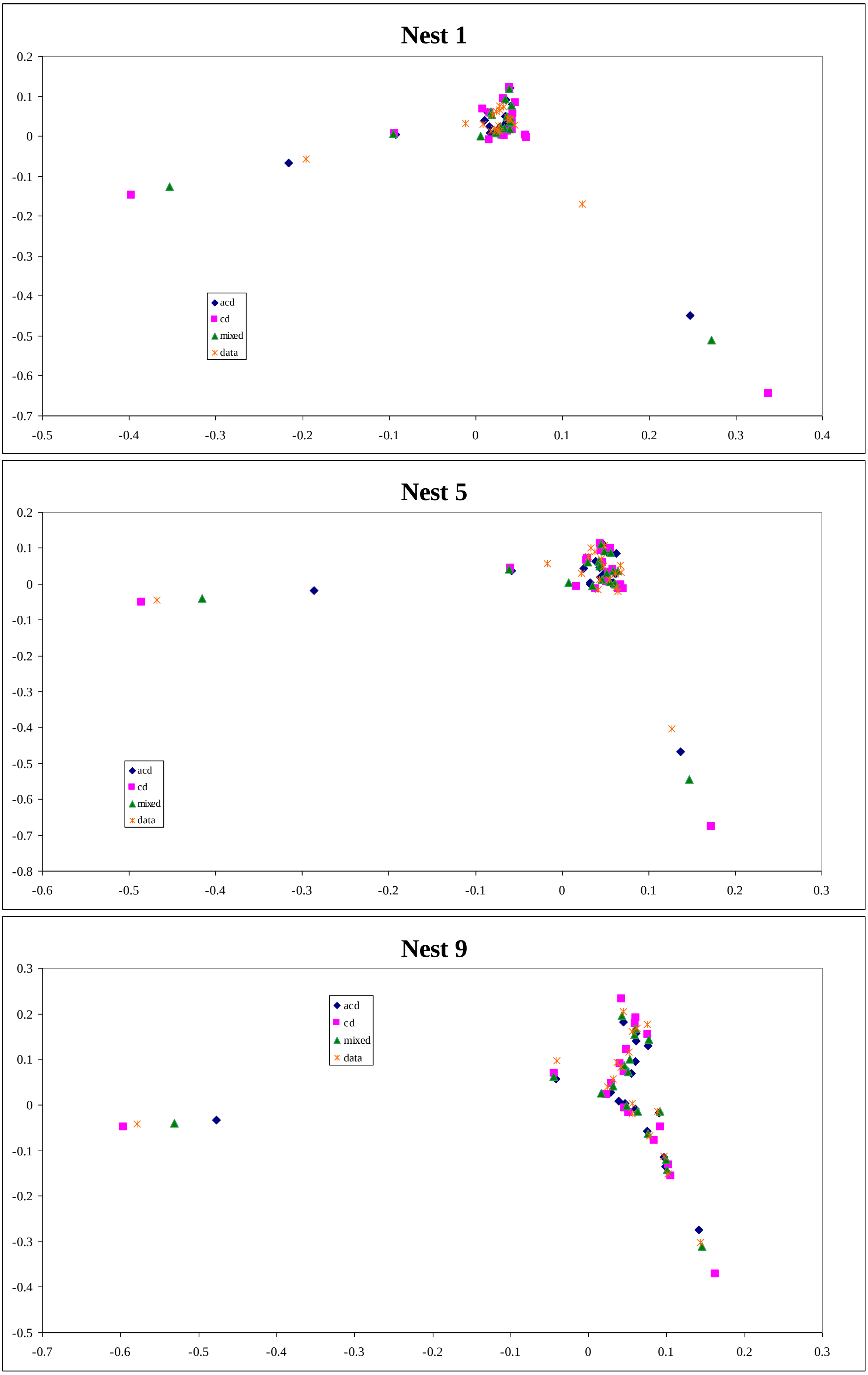


Figure 65 ADAS data Correspondence analysis dimensions 1 and 2: Values predicted from frequency area curves fitted to nests 10 to 16

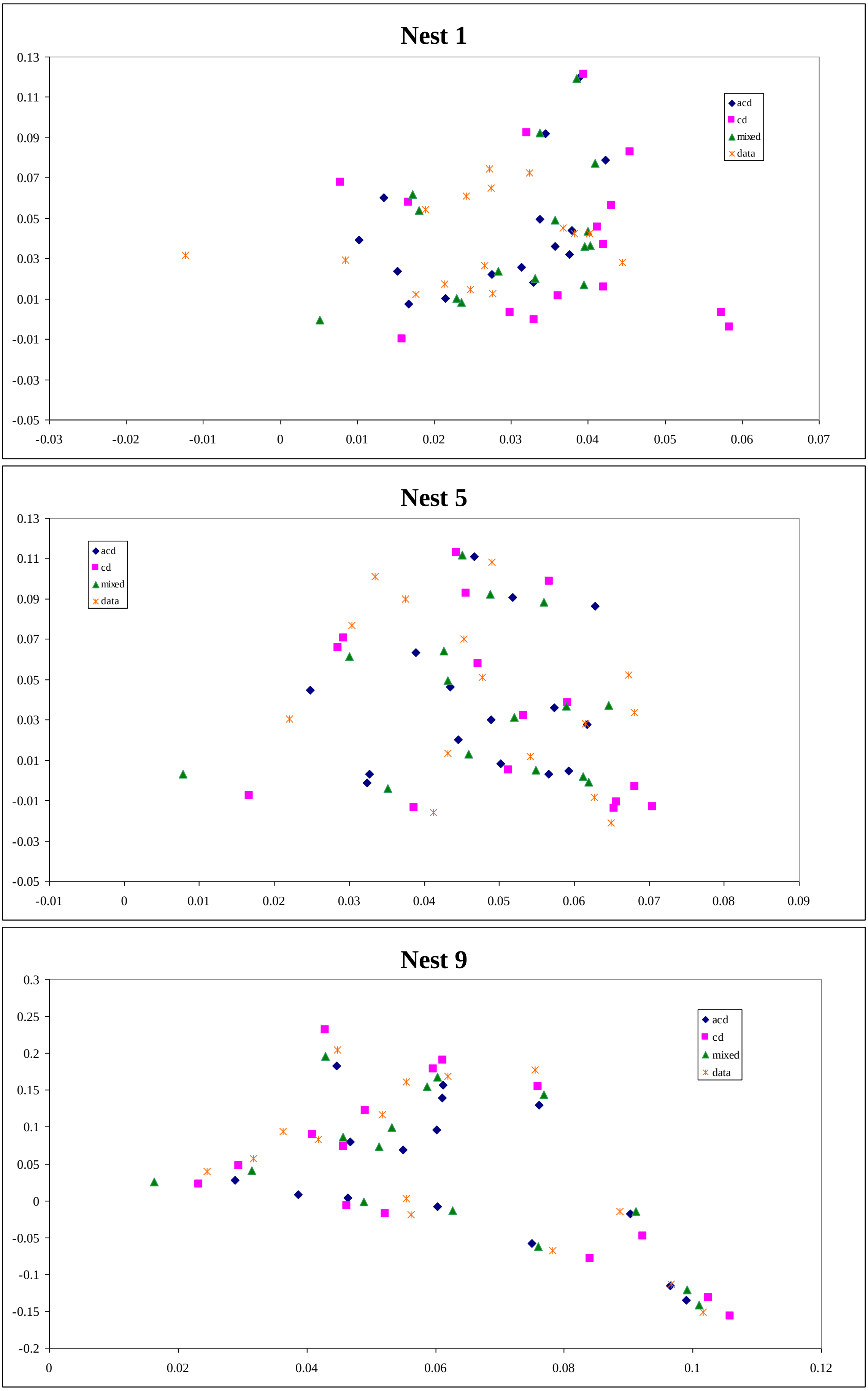


Figure 66 ADAS data Correspondence analysis dimensions 1 and 2: Values predicted from frequency area curves fitted to nests 10 to 16 - Expanded versions of plots in Figure 64

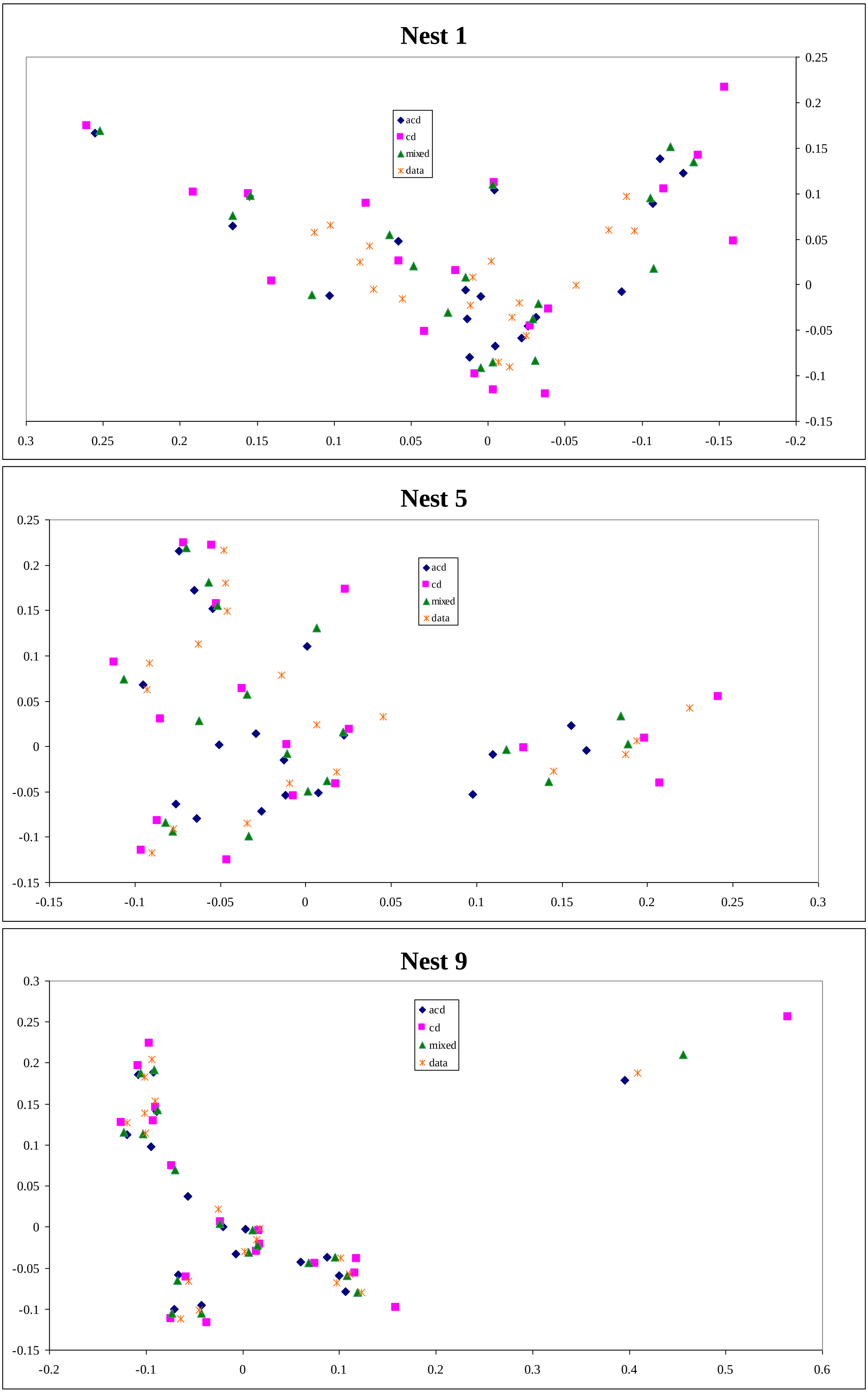


Figure 67 ADAS data Correspondence analysis dimensions 3 and 4: Values predicted from frequency area curves fitted to nests 10 to 16

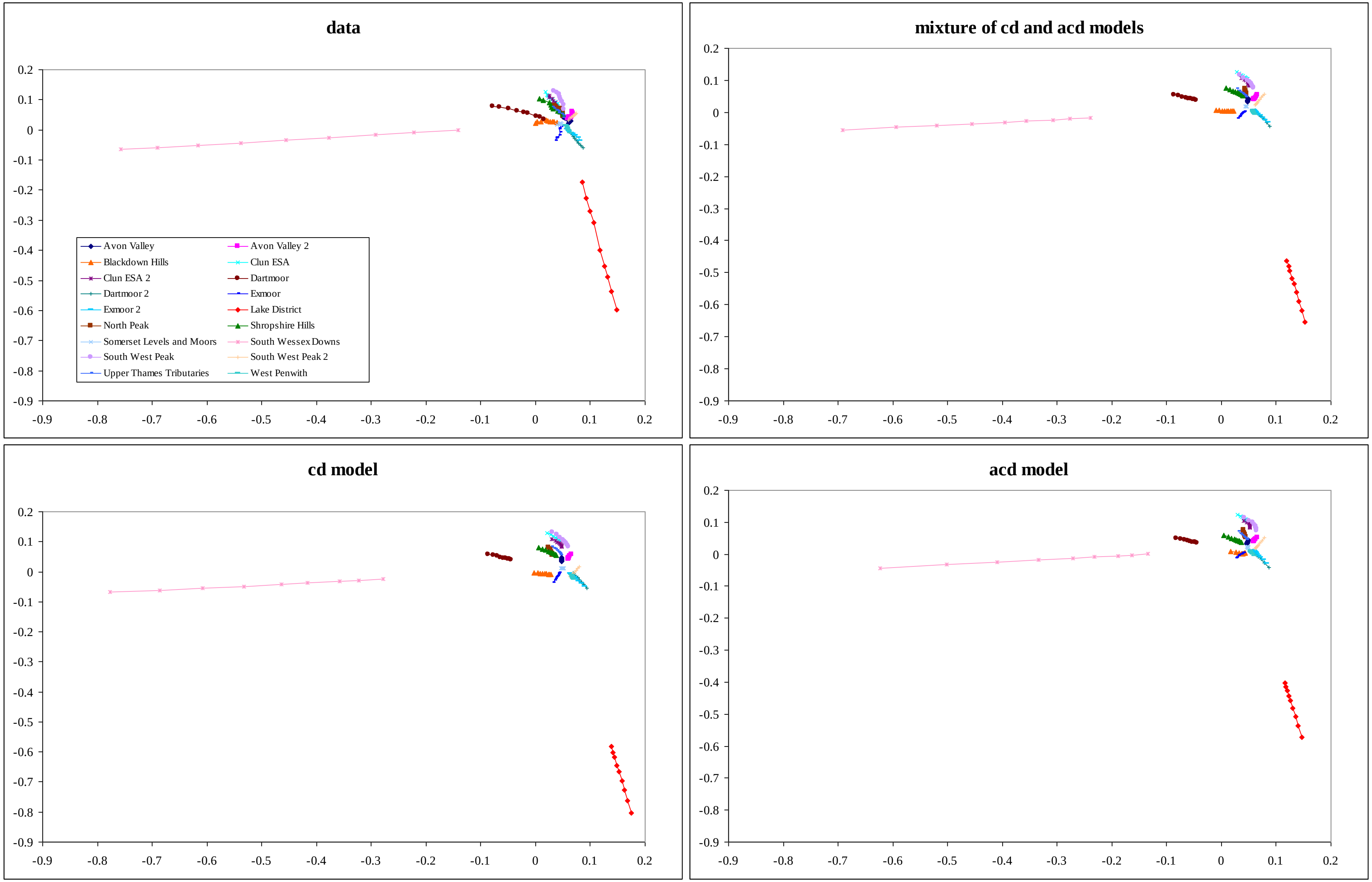


Figure 68 ADAS ESA data correspondence analysis dimensions 1 and 2, nests 1-9 entered as separate points - frequencies predicted from frequency area curves fitted to nests 10-16

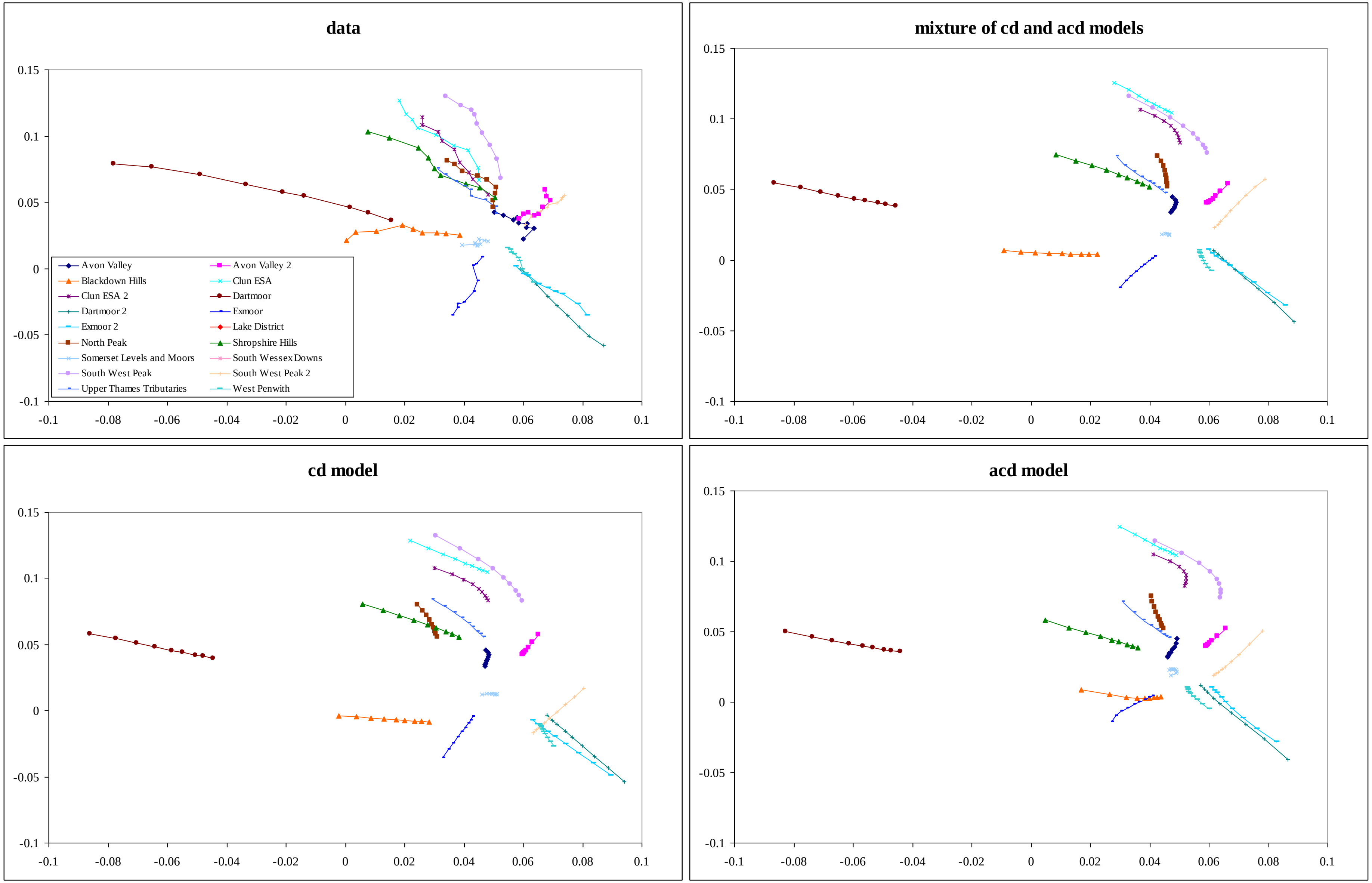


Figure 69 ADAS ESA data correspondence analysis dimensions 1 and 2, nests 1-9 entered as separate points - frequencies predicted from frequency area curves fitted to nests 10-16 - Expanded view of Figure 67

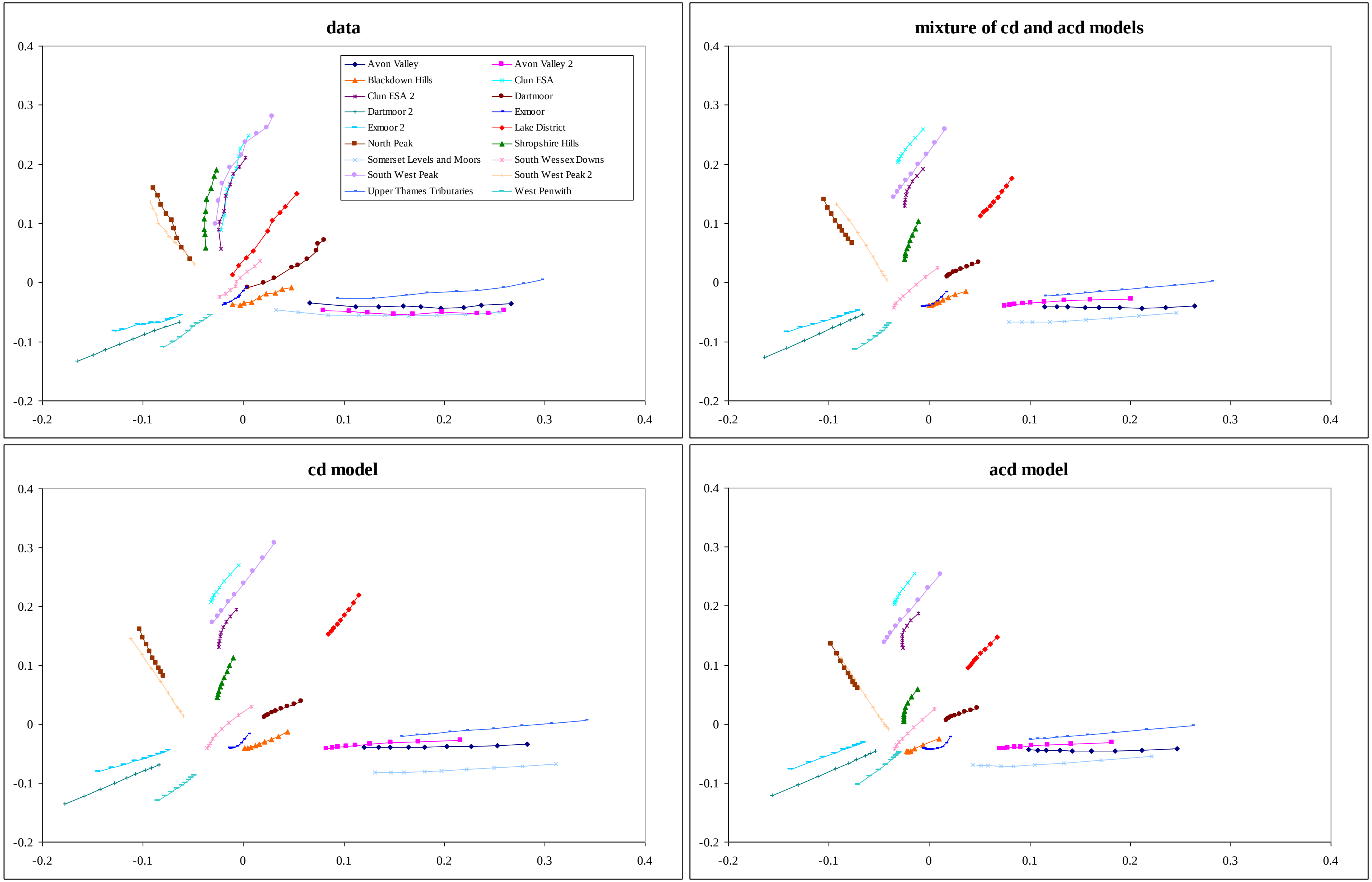


Figure 70 ADAS ESA data correspondence analysis dimensions 3 and 4, nests 1-9 entered as separate points - frequencies predicted from frequency area curves fitted to nests 10-16

7.24 Figures 65 to 70 illustrate the results of using data from nest sizes above 1 m² to predict frequencies at nest sizes 1 m² and below. Figure 65 shows the first two dimensions of the ordination analysis applied to the predicted and actual values at three nest sizes. Each site is represented by four points, one for the data and one for each of the three methods of prediction. For nest size 9 (1 m²) the three methods produce similar results, close to the data points. For smaller nest sizes this is less true, though the concentration of the majority of points in one place makes this difficult to observe. Figure 66 shows expanded views of the plots in Figure 65. For nest 9 it is clear that all method are producing predictions close to the data and the four points representing each site are clustered closely together. This is not true for the smaller nest sizes. At nest size 5 (0.06 m²) some sites can be seen to be represented by clusters of sites but others are not while at nest size 1 (0.004 m²) almost no association can be seen at all. Furthermore, particularly at nest size 1, the data points seem to be closer together than the predicted values, suggesting that the predicted values are more variable. Figure 67 shows the third and fourth dimensions for the same three nest sizes. The same patterns are apparent. Predictions are good for nest size 9 but progressively worse as predictions are made for nest sizes further from the data used for prediction. The smaller variability at nest size 1 is even more pronounced than in Figure 66.

7.25 Figure 68 shows an alternative way of comparing predictions. In this figure the different prediction methods and the data are shown as separate plots with all nest sizes entered onto each plot. A single ordination analysis was used to obtain these plots, however, so that positions are directly comparable across the plots. The plots look relatively similar showing that all prediction methods pick out the main features of the data, at least at the larger nest sizes. It is noticeable, however, that the distance between nest sizes is smaller for the predicted values than it is in the observed data, as evidenced by the shortness of the tracks representing each site. Figure 69, an expanded view of Figure 68, confirms this impression. Predicted values, for all three methods, underestimate the differences between nest sizes at individual sites. A similar tendency is also evident in dimensions three and four (Figure 70).

7.26 Figures 71 to 76 show the results of using data from nest sizes 1 m² and below to predict frequencies at nest sizes above 1 m². The first two dimensions of the ordination results from the individual nest sizes (Figure 71 and expanded version in Figure 72) show that the predictions to larger nest sizes better match the data than when predicting from large to small nests. The points representing each site can be seen to be clustered even at the largest nest size. There is a tendency, however, for the predicted values to be less variable that the data values and this is especially so for predictions made using the *acd* parameterisation. The same effect can be seen in dimensions 3 and 4 (Figure 73). This effect can be seen more clearly when all nests are analysed together (Figures 74 to 76). In these plots it is noticeable that the differences between nest sizes in the data are greater than those in the predicted values. Again the effect is particularly noticeable for the *acd* parameterisation.

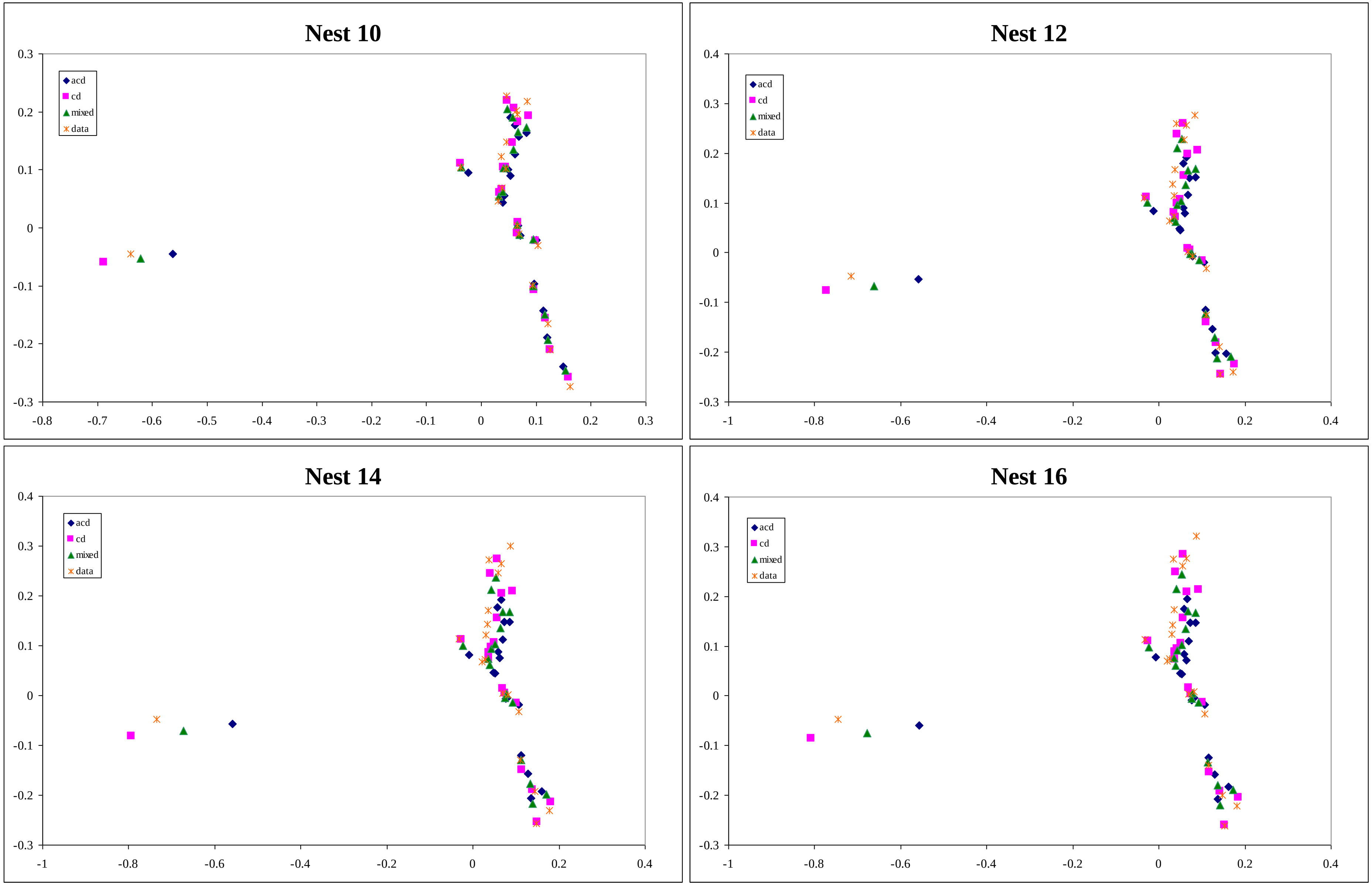


Figure 71 ADAS data Correspondence analysis dimensions 1 and 2, values predicted from frequency area curves fitted to nests 1 to 9

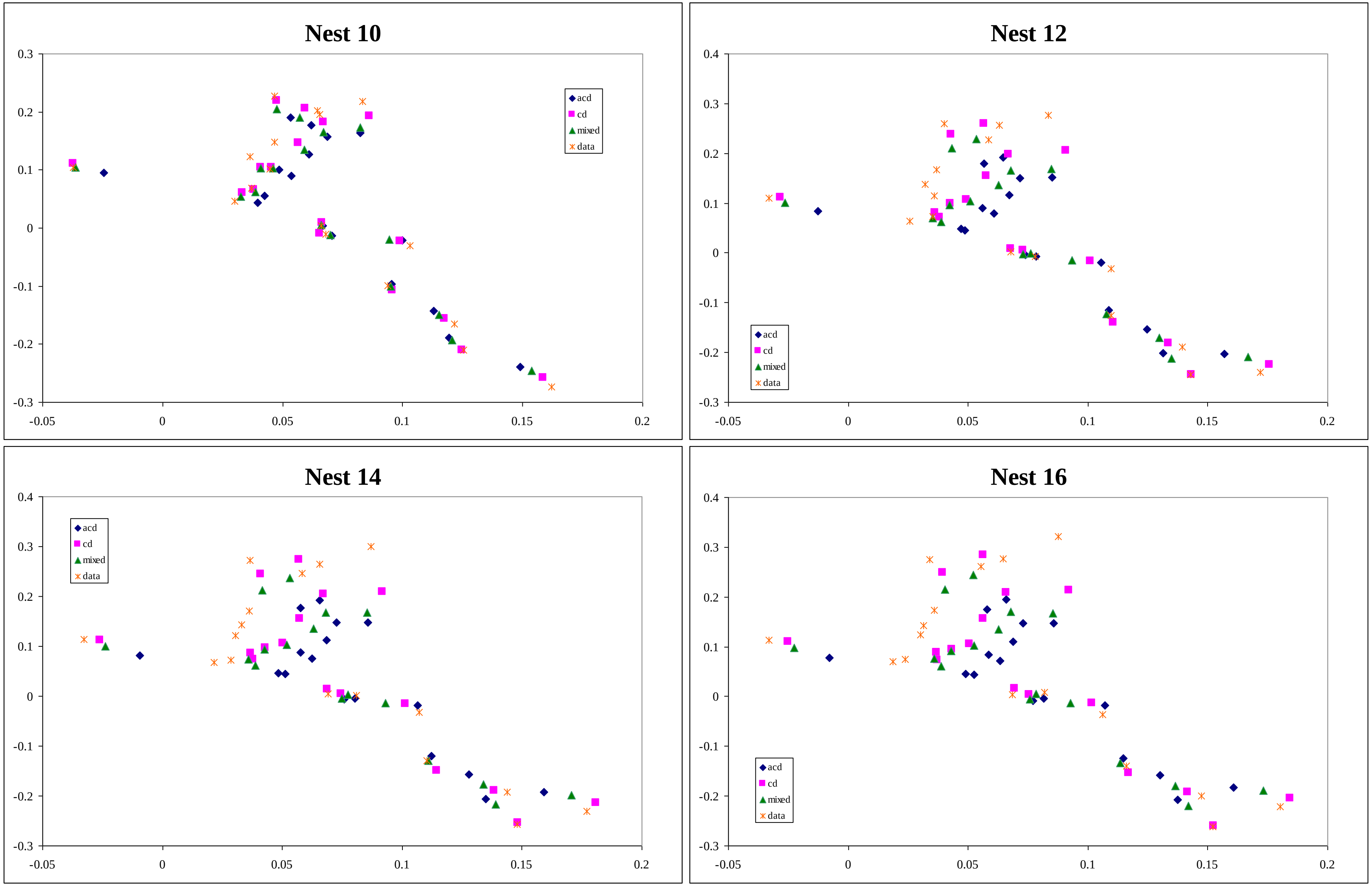


Figure 72 ADAS data Correspondence analysis dimensions 1 and 2, values predicted from frequency area curves fitted to nests 1 to 9 - Expanded versions of plots in Figure 70

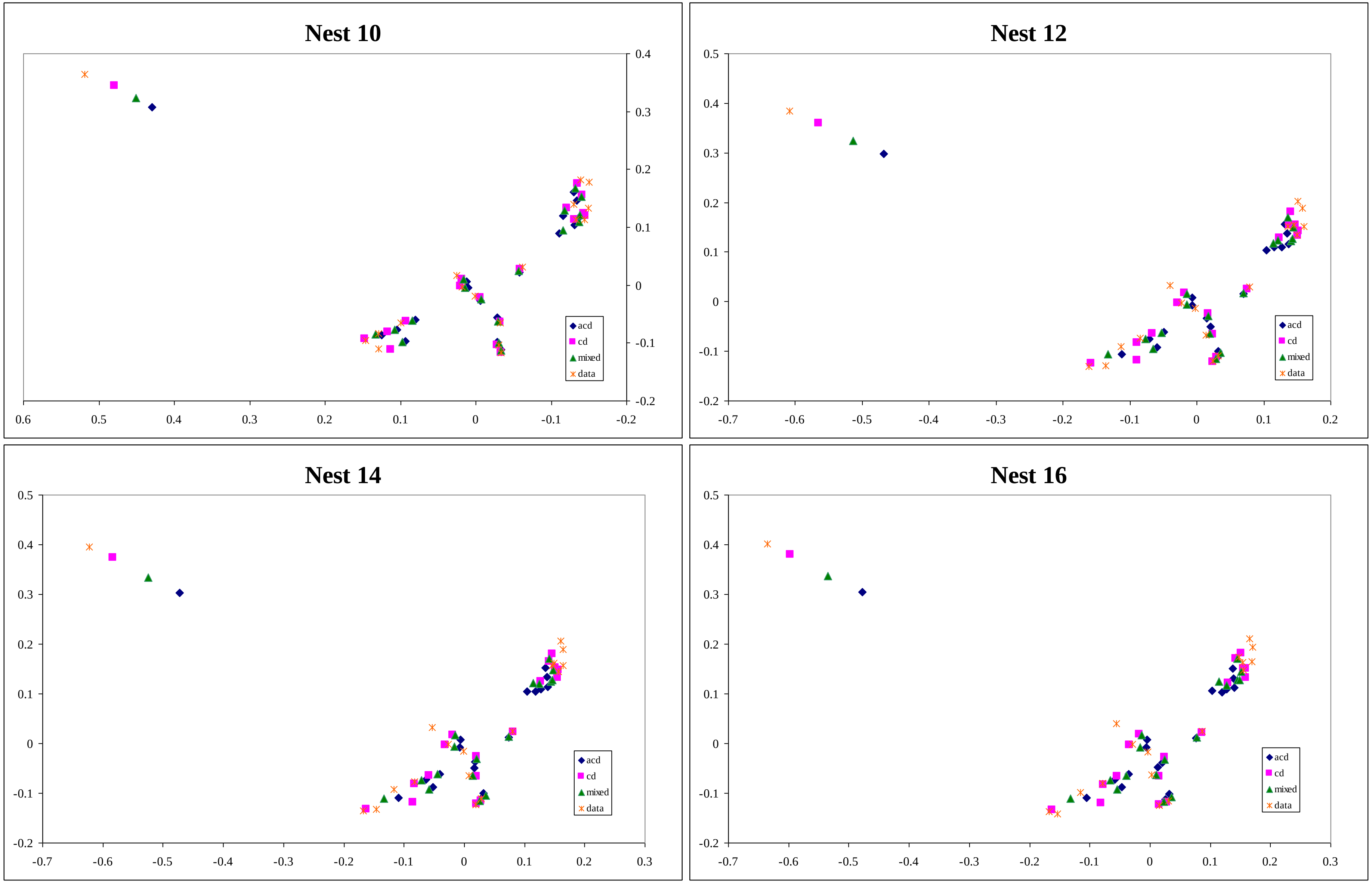


Figure 73 ADAS data Correspondence analysis dimensions 3 and 4, values predicted from frequency area curves fitted to nests 1 to 9

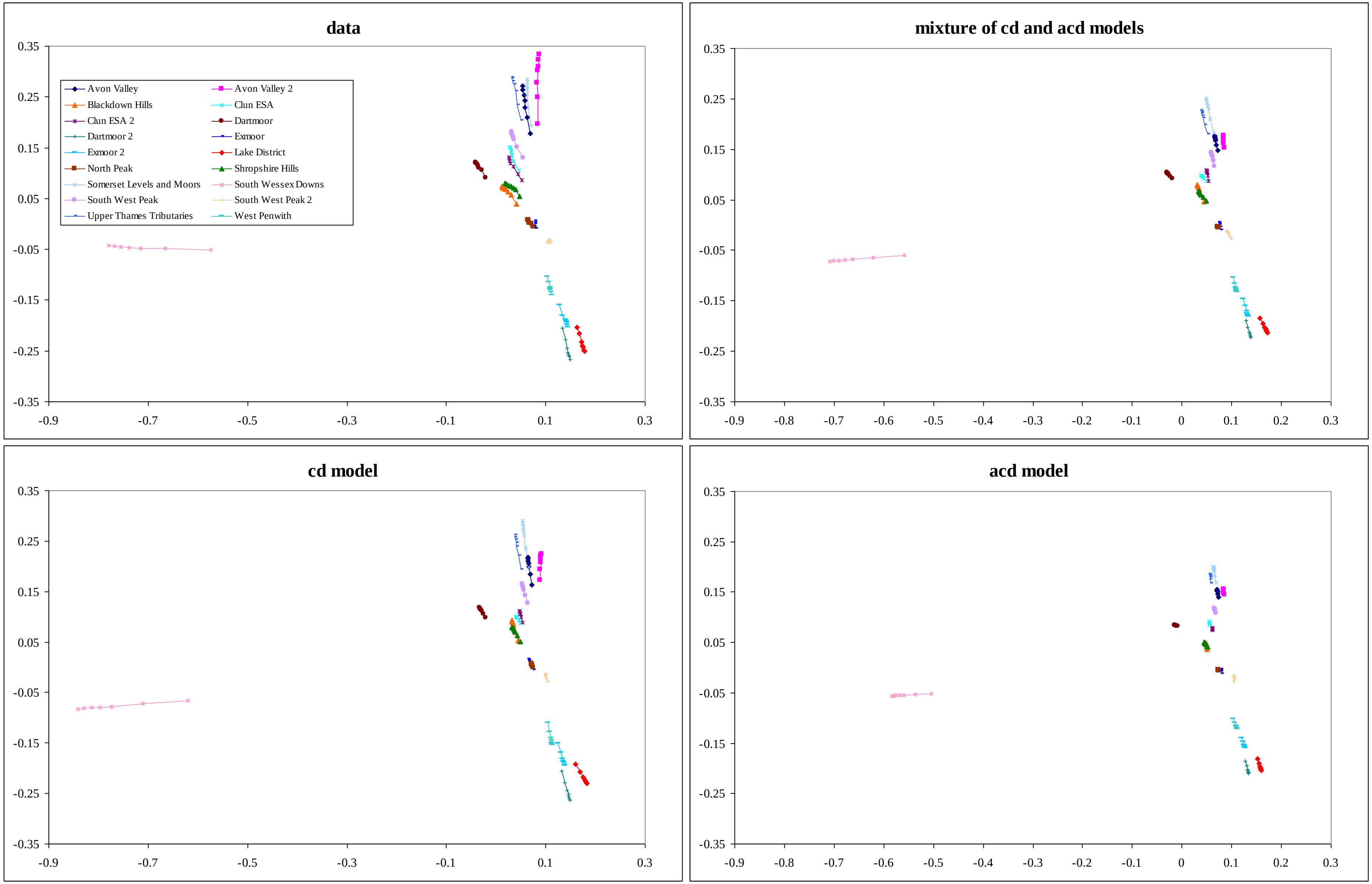


Figure 74 ADAS ESA data correspondence analysis dimensions 1 and 2, nests 10-16 entered as separate points - frequencies predicted from frequency area curves fitted to nests 1-9

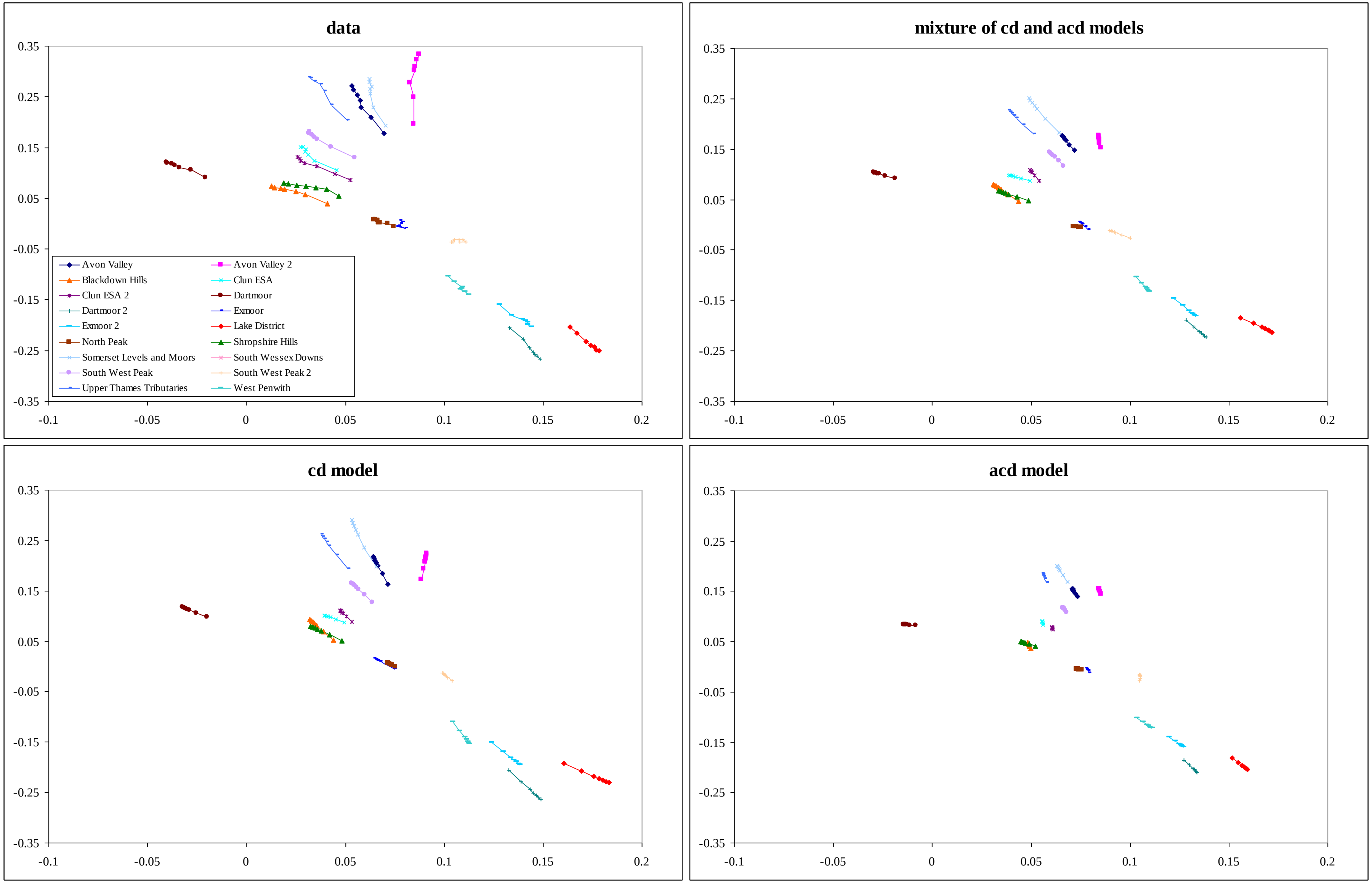


Figure 75 ADAS ESA data correspondence analysis dimensions 1 and 2, nests 10-16 entered as separate points - frequencies predicted from frequency area curves fitted to nests 1-9 - Expanded view of Figure 73

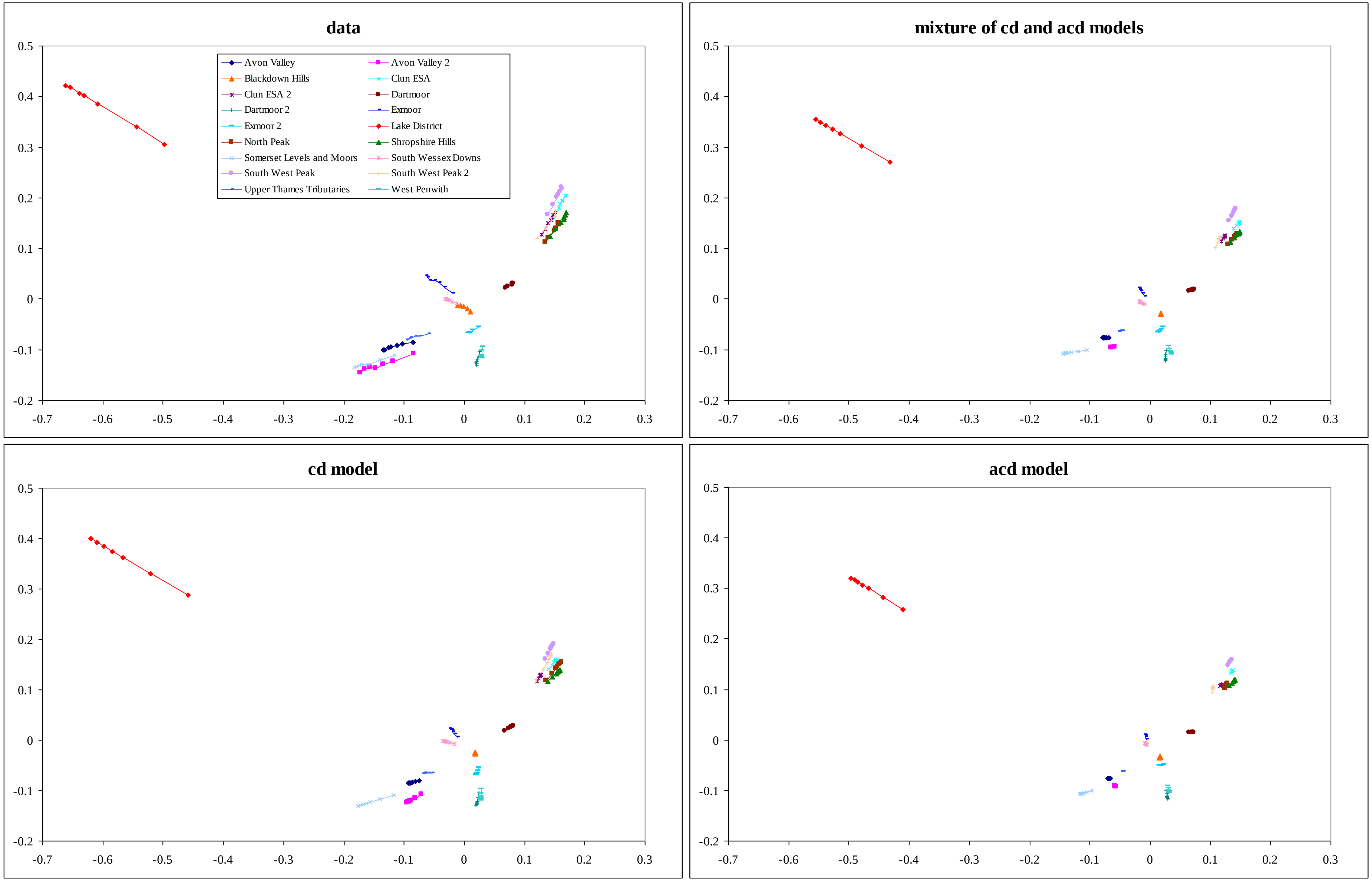


Figure 76 ADAS ESA data correspondence analysis dimensions 3 and 4, nests 10-16 entered as separate points - frequencies predicted from frequency area curves fitted to nests 1-9

Accuracy of standardised datasets

7.27 As shown above standardised datasets produced using fitted frequency-area curves are affected by the inaccuracies of the extrapolation process in addition to any inherent sampling variation. The magnitude of these inaccuracies is examined in this section.

7.28 Figure 77 shows, for the ADAS dataset, the average frequencies at each nest size for the data and the predictions from the frequency-area curves fitted to the data for nests greater than 1 m^2. Predictions are shown for the *cd* parameterisation, the *acd* parameterisation, and a mixture of the *cd* parameterisation for the majority of species and the *acd* parameterisation for those species for which it is significantly better. All three prediction methods give values close to the data at the larger nest sizes, from which the prediction curves were derived, but appear to fit less well at smaller nest sizes, although the closeness of the plotted points makes it difficult to observe the size of the discrepancies. Figure 78 plots the same values using logarithmic scales on both axes. The discrepancies are much clearer. The *cd* parameterisation fits the data well except at the three smallest nest sizes. In contrast the *acd* parameterisation underestimates the average frequency at all nest sizes except those used for curve fitting. The mixed approach is intermediate between the other two methods better fitting than the *acd* method but fitting less well than the *cd* parameterisation. None of the prediction methods accurately models the increasing drop-off shown by the data at the smallest datasets.

7.29 Figures 77 and 78 show cumulative frequencies whereas the curves are fitted to the data in the form in which it is collected, that is, the additional frequency found at each nest size. Figure 79 shows the data and predicted values in this form. An additional final point has been added to this plot showing the average of the proportion of plots in which each species is not found. As before all prediction methods fit well to the larger nest sizes from which they were derived but the fit at the smaller nest sizes is unclear. Figure 80 shows the same data using logarithmic scales. It can now be seen that all prediction methods are underestimating the frequencies at each of the five smaller nest sizes except the first.

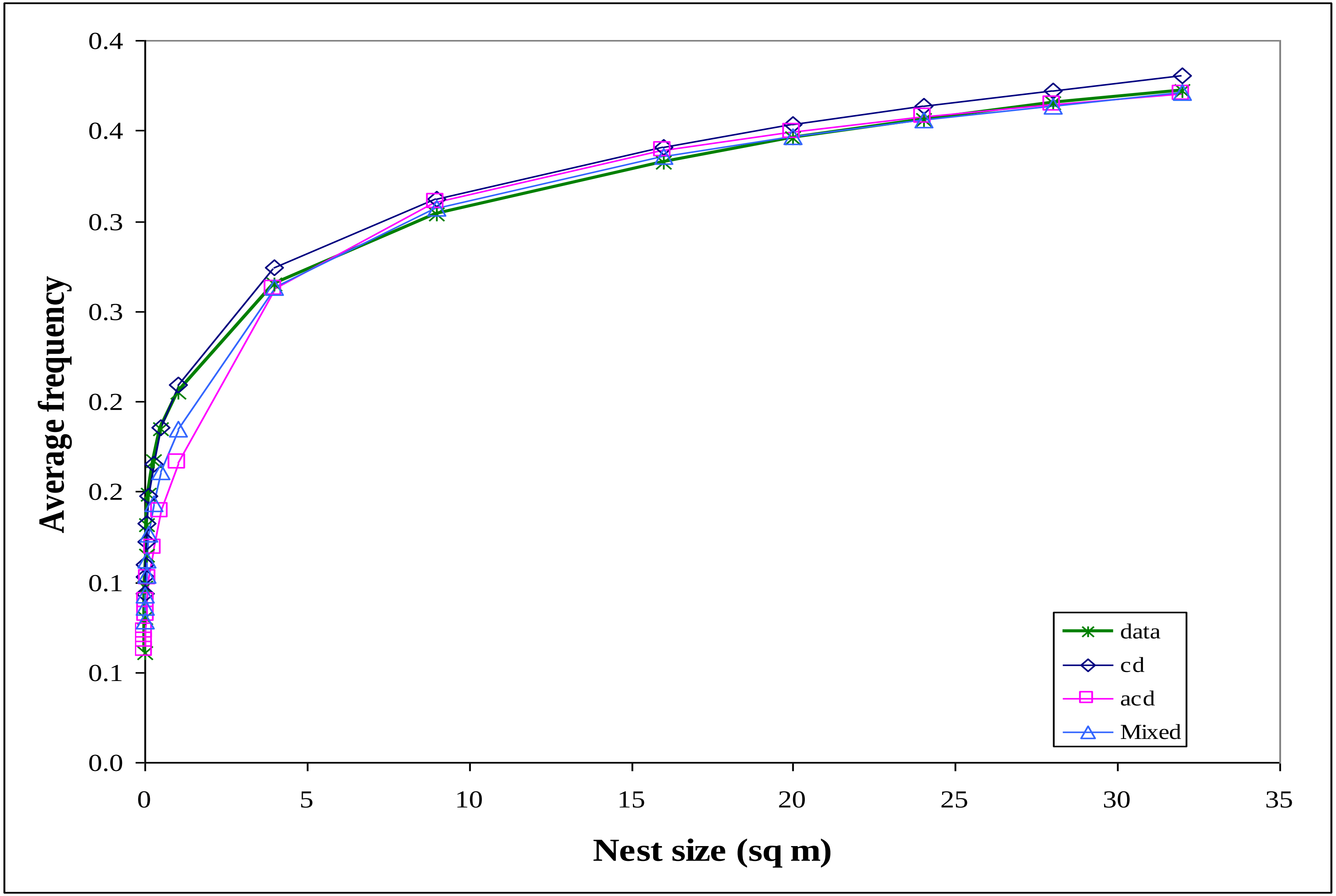


Figure 77 ADAS ESA data - Cumulative frequencies predicted from frequency area curves fitted to nests 10-16

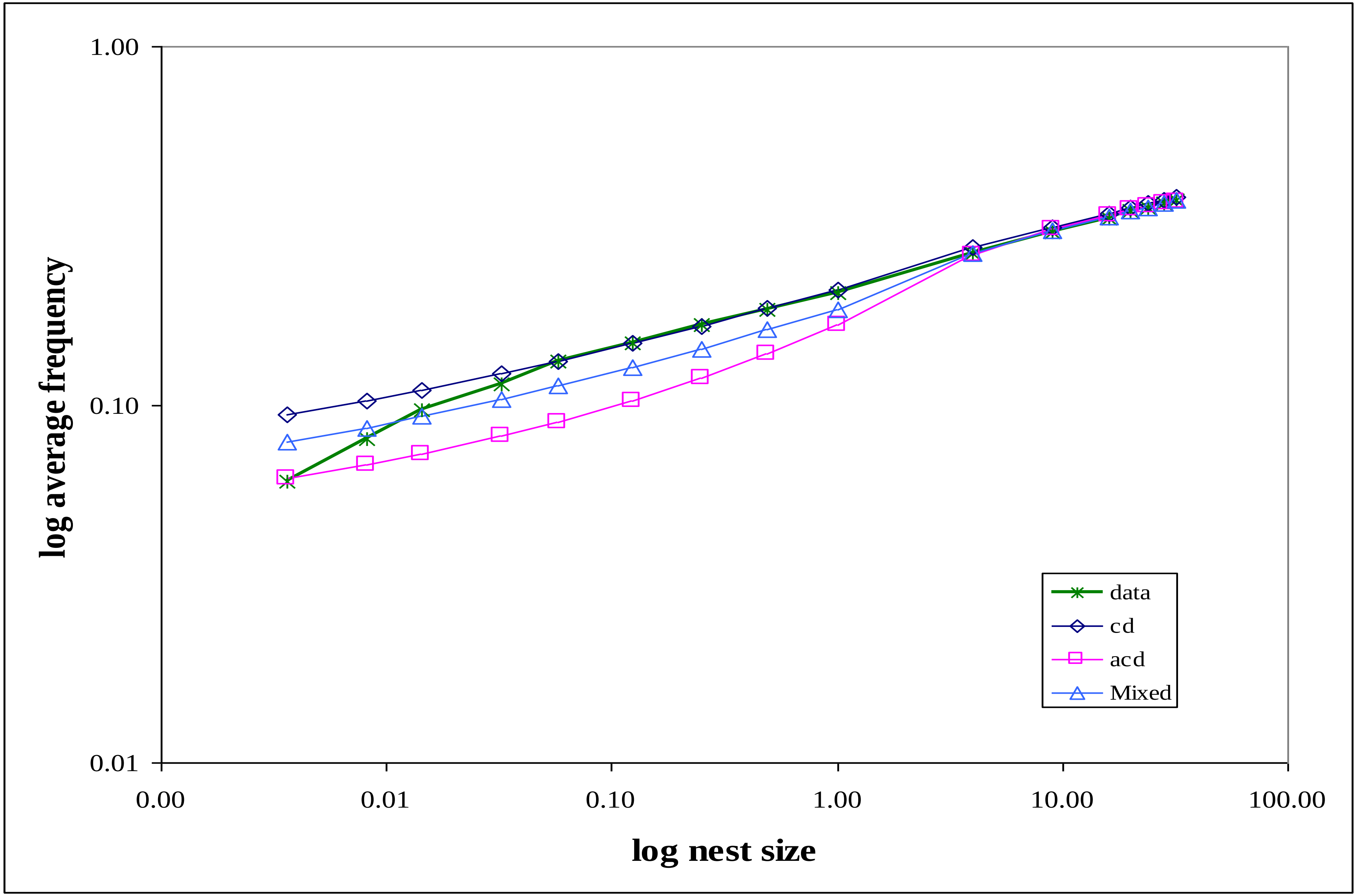


Figure 78 ADAS ESA data - Figure 77 with logarithmic scales

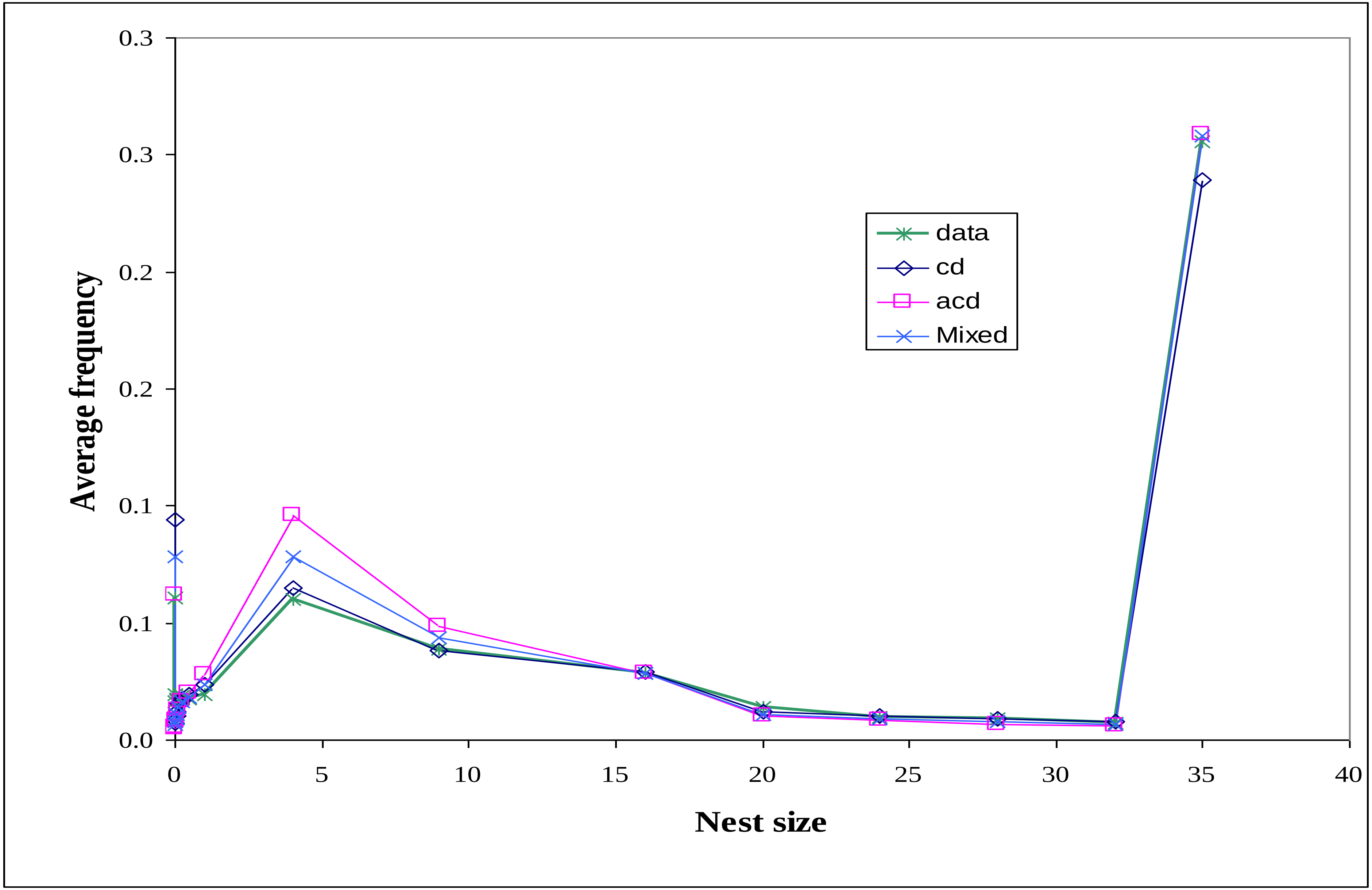


Figure 79 ADAS ESA data - Frequencies predicted from frequency area curves fitted to nests 10-16

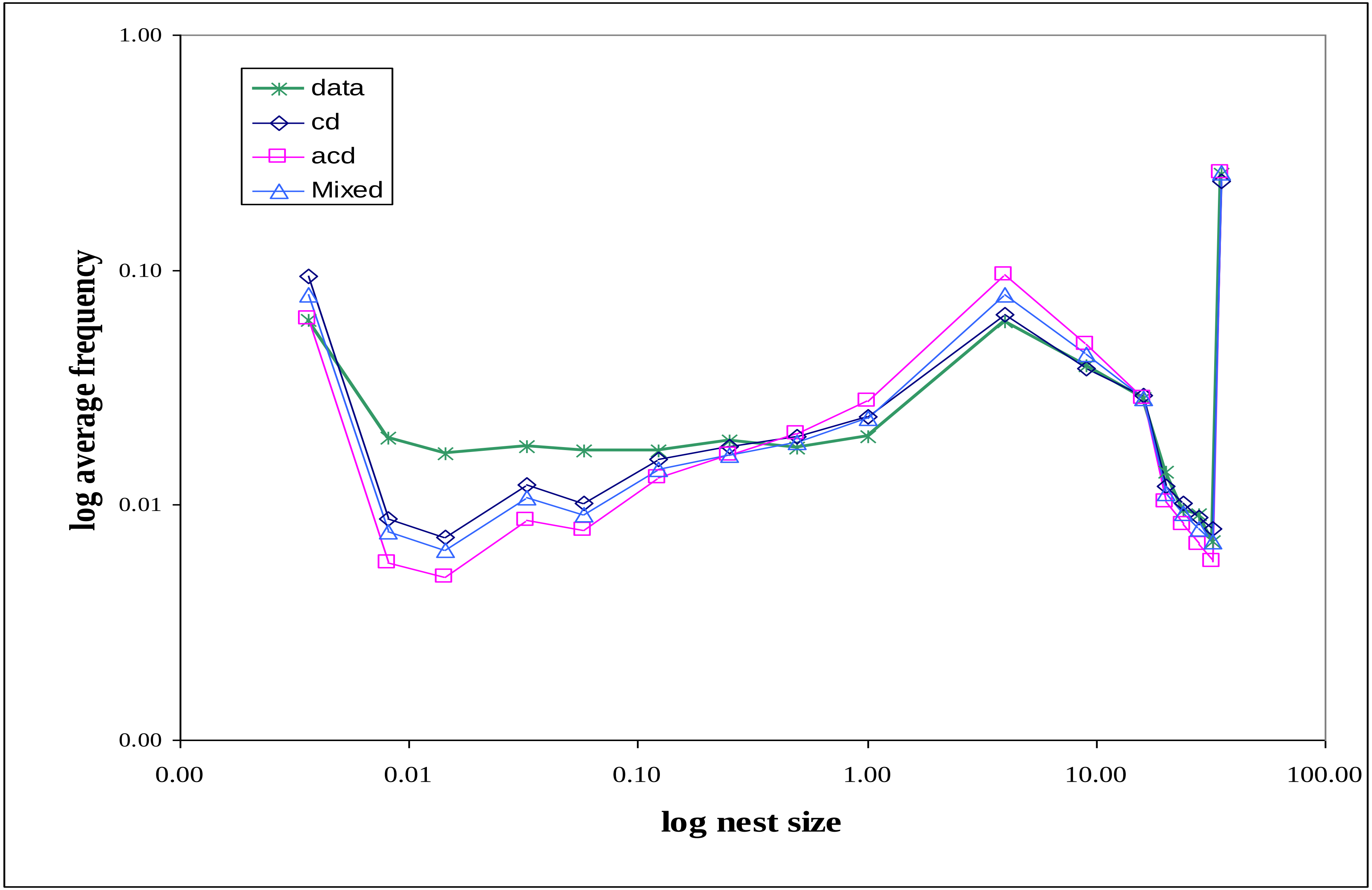


Figure 80 ADAS ESA data - Figure 79 with logarithmic scale

7.30 The average cumulative frequencies for each nest, for the data and each prediction method, are given in Table 37, broken down by the amount of data observed for each species. Detailed comparison of the values shows that the prediction methods perform worst when the amount of observed data is small, that is, for species found in less than ten quadrats, and for the two smallest nest sizes. Table 38 gives the ratios of the predicted frequencies to the data frequencies, clarifying the extent of these errors. It also shows that in the other cells of the table all prediction methods somewhat underestimate frequencies, particularly the *acd* parameterisation.

Table 37 Average frequency for data and predicted values by nest size and number of times species observed: Predictions from curves fitted to data from nests above 1 m²

	Number of quadrats species found in	Nest size (sq m)								
		0.004	**0.008**	**0.014**	**0.032**	**0.058**	**0.123**	**0.25**	**0.49**	**1**
Data	3-4	0.005	0.008	0.009	0.012	0.014	0.017	0.020	0.024	0.031
	5-9	0.011	0.017	0.023	0.030	0.038	0.045	0.056	0.065	0.076
	10-14	0.042	0.058	0.073	0.090	0.108	0.124	0.142	0.160	0.186
	15-19	0.109	0.139	0.167	0.199	0.229	0.259	0.293	0.324	0.355
	20-29	0.133	0.175	0.205	0.240	0.269	0.300	0.331	0.363	0.395
	30-39	0.186	0.237	0.280	0.317	0.354	0.387	0.423	0.455	0.481
	40-49	0.167	0.196	0.222	0.256	0.280	0.318	0.347	0.376	0.401
	50-99	0.156	0.214	0.265	0.314	0.360	0.407	0.456	0.501	0.546
	all	0.061	0.080	0.097	0.115	0.132	0.149	0.167	0.185	0.205
cd	3-4	0.047	0.049	0.050	0.051	0.053	0.055	0.058	0.062	0.066
	5-9	0.041	0.045	0.048	0.053	0.057	0.064	0.072	0.081	0.094
	10-14	0.072	0.080	0.086	0.096	0.106	0.121	0.139	0.160	0.187
	15-19	0.171	0.181	0.190	0.205	0.218	0.237	0.261	0.288	0.323
	20-29	0.163	0.180	0.195	0.220	0.240	0.271	0.304	0.339	0.379
	30-39	0.194	0.216	0.235	0.265	0.289	0.326	0.366	0.407	0.454
	40-49	0.176	0.192	0.205	0.226	0.243	0.270	0.302	0.336	0.379
	50-99	0.174	0.201	0.224	0.262	0.294	0.342	0.395	0.450	0.514
	all	0.094	0.103	0.110	0.122	0.132	0.148	0.166	0.185	0.209

Table continued…

	Number of quadrats species found in	Nest size (sq m)								
		0.004	**0.008**	**0.014**	**0.032**	**0.058**	**0.123**	**0.25**	**0.49**	**1**
acd	3-4	0.016	0.016	0.017	0.018	0.019	0.020	0.022	0.024	0.027
	5-9	0.024	0.026	0.028	0.031	0.034	0.039	0.045	0.052	0.064
	10-14	0.057	0.062	0.066	0.073	0.080	0.091	0.106	0.125	0.152
	15-19	0.152	0.159	0.166	0.177	0.187	0.203	0.223	0.250	0.288
	20-29	0.097	0.107	0.116	0.133	0.149	0.175	0.208	0.249	0.303
	30-39	0.133	0.148	0.160	0.182	0.201	0.234	0.274	0.322	0.386
	40-49	0.117	0.128	0.138	0.156	0.172	0.199	0.232	0.272	0.325
	50-99	0.124	0.143	0.160	0.190	0.216	0.261	0.315	0.380	0.462
	all	0.062	0.068	0.073	0.082	0.089	0.102	0.119	0.139	0.167
Mixed	3-4	0.016	0.017	0.017	0.019	0.020	0.022	0.024	0.026	0.030
	5-9	0.032	0.035	0.038	0.042	0.046	0.052	0.059	0.067	0.078
	10-14	0.067	0.073	0.079	0.088	0.096	0.110	0.126	0.145	0.170
	15-19	0.163	0.172	0.180	0.193	0.204	0.222	0.244	0.269	0.304
	20-29	0.148	0.164	0.177	0.200	0.218	0.247	0.279	0.312	0.352
	30-39	0.178	0.197	0.212	0.238	0.260	0.292	0.328	0.368	0.418
	40-49	0.176	0.192	0.205	0.226	0.243	0.270	0.302	0.336	0.379
	50-99	0.152	0.177	0.197	0.232	0.261	0.306	0.356	0.413	0.483
	all	0.079	0.086	0.093	0.103	0.113	0.127	0.143	0.162	0.185

Table 38 Ratio of predicted average frequency to average frequency for data by nest size and number of times species observed: Predictions from curves fitted to data from nests above 1 m^2

	Number of quadrats species found in	Nest size (sq m)								
		0.004	**0.008**	**0.014**	**0.032**	**0.058**	**0.123**	**0.25**	**0.49**	**1**
cd	3-4	8.94	6.46	5.23	4.46	3.75	3.19	2.84	2.61	2.16
	5-9	3.65	2.57	2.03	1.76	1.52	1.44	1.30	1.25	1.24
	10-14	1.73	1.36	1.18	1.07	0.98	0.98	0.98	1.00	1.01
	15-19	1.57	1.31	1.14	1.03	0.95	0.92	0.89	0.89	0.91
	20-29	1.22	1.03	0.95	0.91	0.89	0.90	0.92	0.93	0.96
	30-39	1.04	0.91	0.84	0.84	0.82	0.84	0.87	0.90	0.94
	40-49	1.05	0.98	0.92	0.88	0.87	0.85	0.87	0.90	0.94
	50-99	1.12	0.94	0.85	0.83	0.82	0.84	0.87	0.90	0.94
	all	1.54	1.28	1.13	1.06	1.01	1.00	0.99	1.00	1.02
acd	3-4	2.99	2.19	1.79	1.56	1.33	1.16	1.07	1.02	0.89
	5-9	2.16	1.52	1.20	1.04	0.91	0.87	0.80	0.80	0.84
	10-14	1.36	1.06	0.90	0.81	0.74	0.73	0.74	0.78	0.82
	15-19	1.40	1.15	0.99	0.89	0.81	0.78	0.76	0.77	0.81
	20-29	0.72	0.61	0.57	0.55	0.55	0.58	0.63	0.68	0.77
	30-39	0.72	0.62	0.57	0.58	0.57	0.60	0.65	0.71	0.80
	40-49	0.70	0.65	0.62	0.61	0.61	0.63	0.67	0.72	0.81
	50-99	0.79	0.67	0.60	0.60	0.60	0.64	0.69	0.76	0.85
	all	1.02	0.85	0.75	0.71	0.68	0.69	0.71	0.75	0.81
mixed	3-4	3.01	2.23	1.84	1.62	1.40	1.24	1.16	1.11	0.97
	5-9	2.85	2.03	1.61	1.41	1.22	1.17	1.06	1.03	1.02
	10-14	1.60	1.25	1.08	0.98	0.89	0.89	0.89	0.91	0.92
	15-19	1.50	1.24	1.08	0.97	0.89	0.86	0.83	0.83	0.86
	20-29	1.11	0.93	0.86	0.83	0.81	0.82	0.84	0.86	0.89
	30-39	0.96	0.83	0.76	0.75	0.73	0.75	0.78	0.81	0.87
	40-49	1.05	0.98	0.92	0.88	0.87	0.85	0.87	0.90	0.94
	50-99	0.98	0.83	0.74	0.74	0.73	0.75	0.78	0.82	0.88
	all	1.29	1.07	0.96	0.90	0.86	0.85	0.85	0.87	0.90

7.31 Figure 81 shows the average cumulative frequencies at each nest size for the data and the predictions from the frequency-area curves fitted to the data for nests 1 m^2 and below. In this case the predicted values at the larger nest sizes clearly differ from the data. The *acd* parameterisation produces a particularly poor fit, substantially underestimating the observed values. Using logarithmic scales (Figure 82) shows the much better fit at the smaller nest sizes, from which the fitted curves were derived, and clarifies the reason for the poor performance of the *acd* parameterisation. The additional parameter allows this parameterisation to approximate more closely to the curvature of the data to which it is fitted but this also increases the curvature at the larger nest sizes to an excessive extent. Using the additional frequencies found at each nest size instead of the cumulative frequencies (Figures 83 and 84) confirms this interpretation.

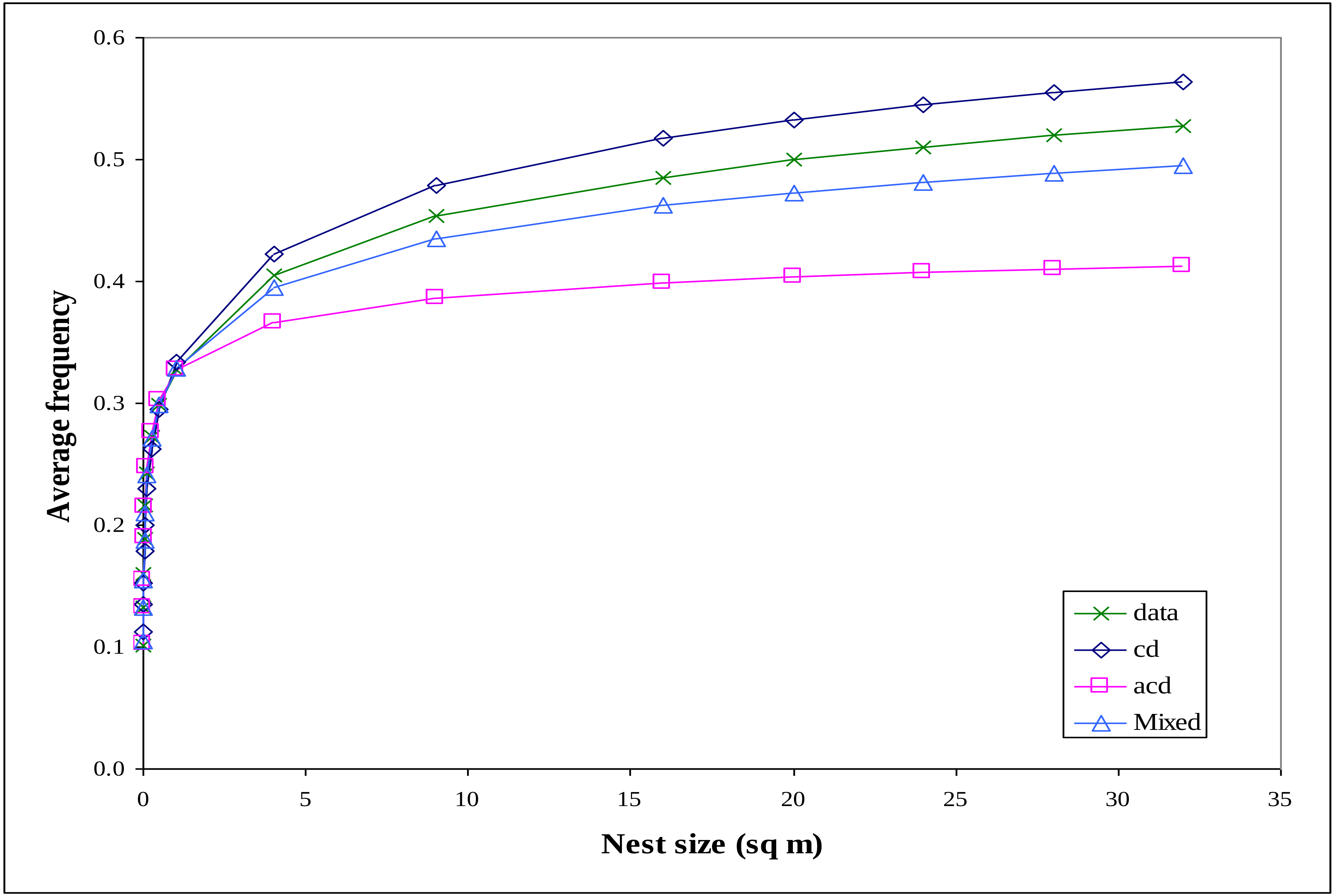


Figure 81 ADAS ESA data - Cumulative frequencies predicted from frequency area curves fitted to nests 1-9

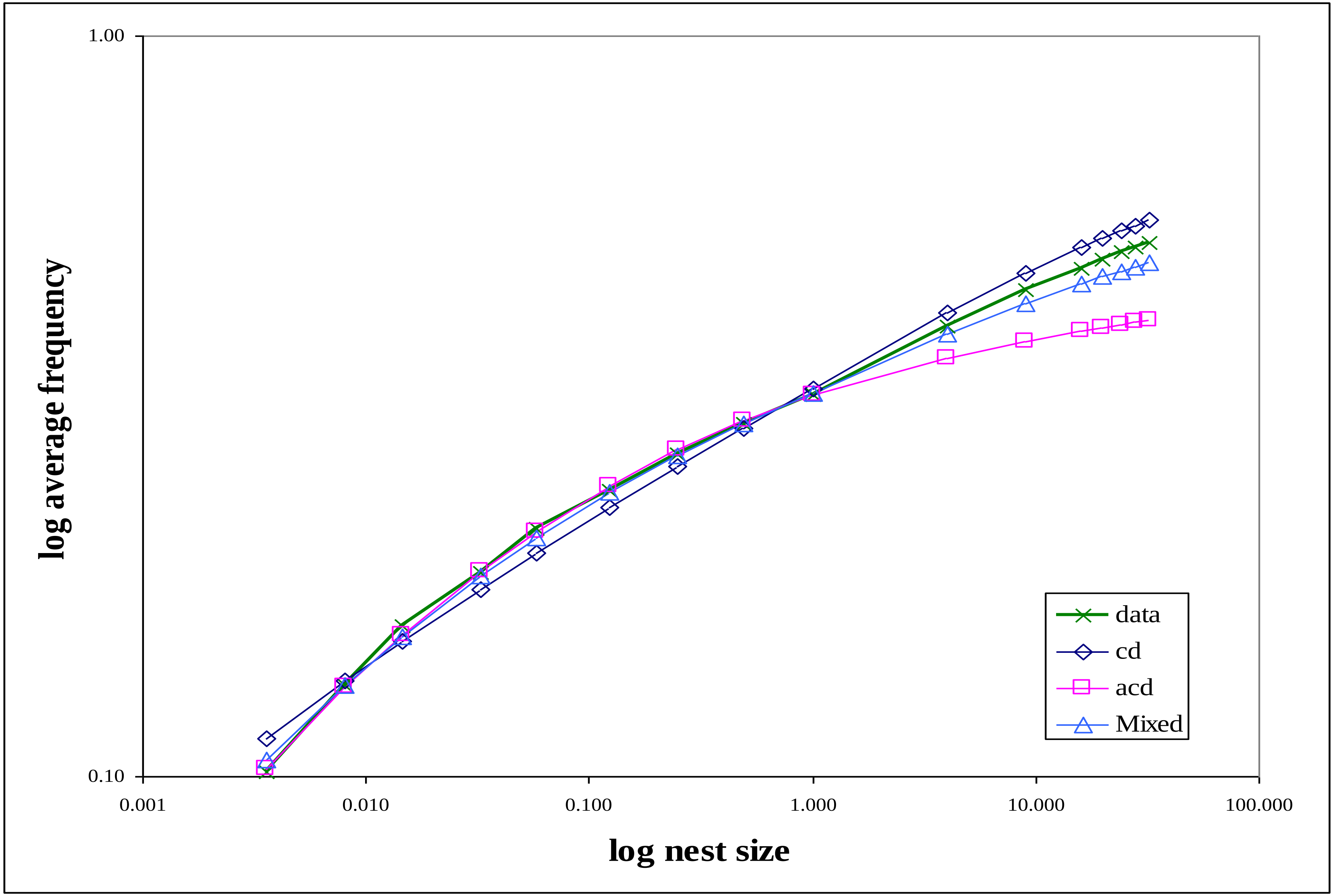


Figure 82 ADAS ESA data - Figure 81 with logarithmic scales

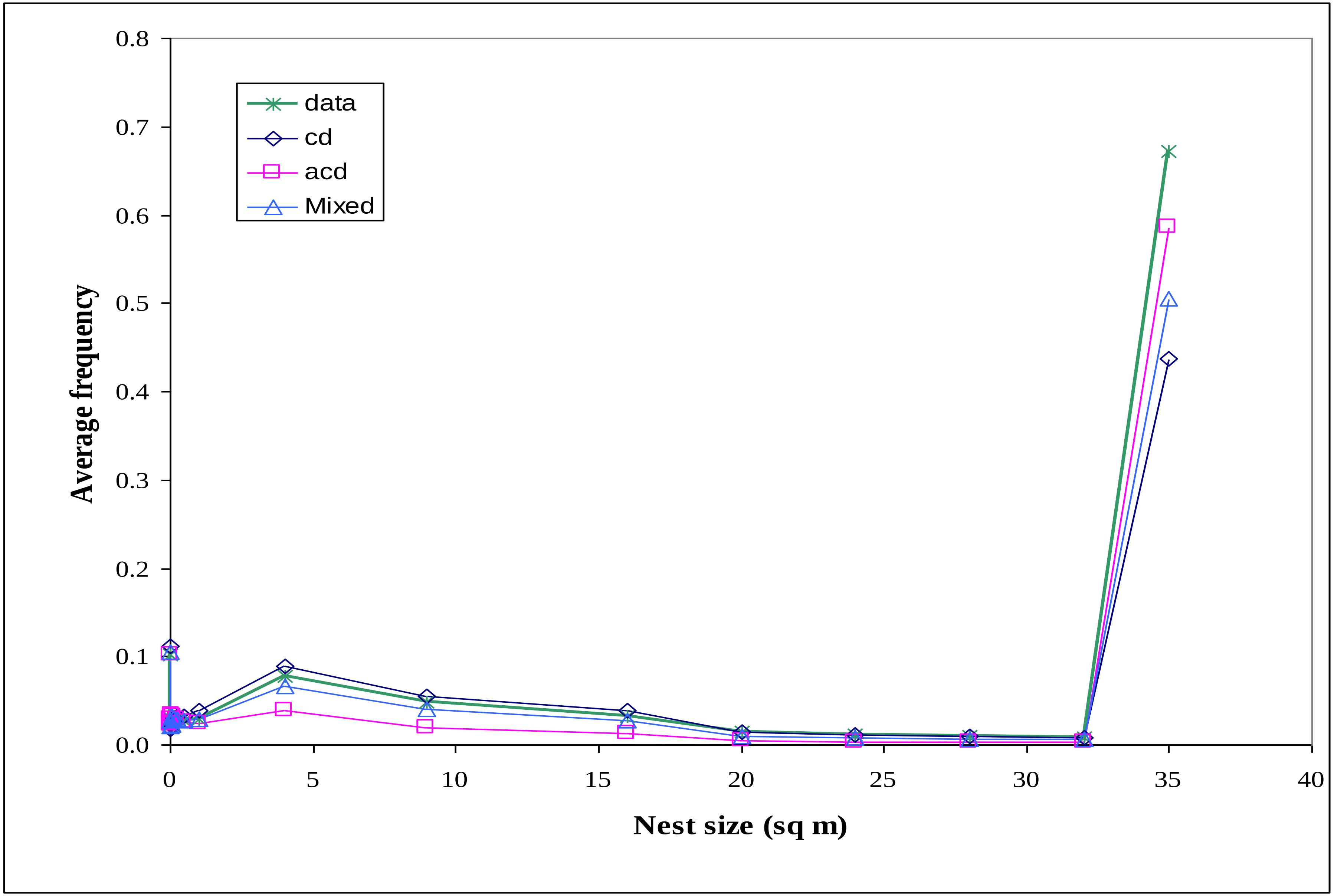


Figure 83 ADAS ESA data - Frequencies predicted from frequency area curves fitted to nests 1-9

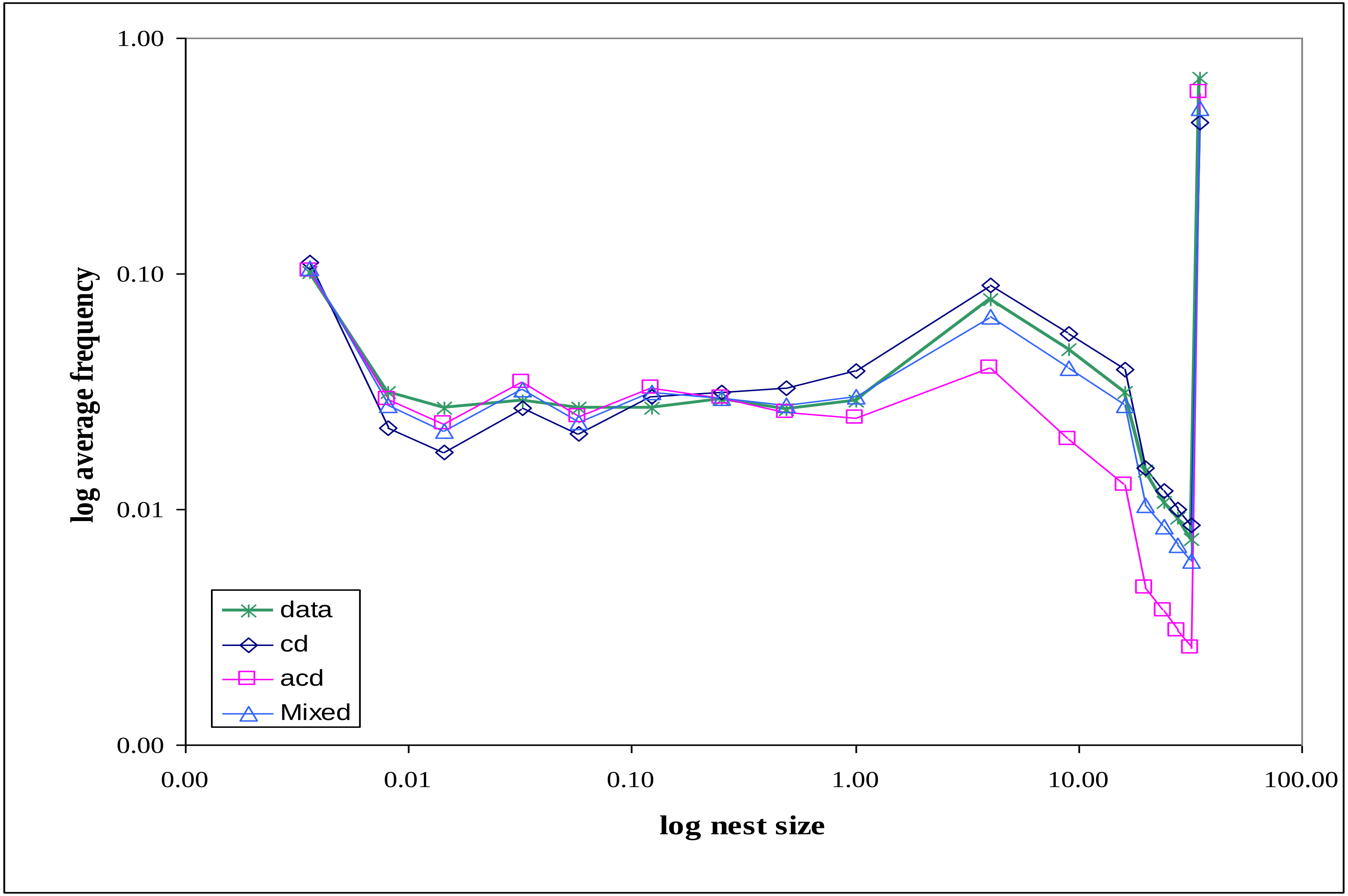


Figure 84 ADAS ESA data - Figure 83 with logarithmic scales

7.32 Table 39 gives the observed and predicted cumulative frequencies broken down by the number of quadrats in which species are found, and Table 40 gives the ratios of the predicted to observed values. As above predictions are poor for species found in less than ten quadrats and the size of the discrepancy increases for nest sizes further from the data to which the curves are fitted. The *acd* parameterisation considerably underestimates the observed frequencies for species found in more than ten quadrats but, unlike the previous case, the *cd* parameterisation does not.

Table 39 Average frequency for data and predicted values by nest size and number of times species observed: Predictions from curves fitted to data from nests 1 m^2 and below

	Number of quadrats species found in	Nest size (sq m)						
		4	**9**	**16**	**20**	**24**	**28**	**32**
Data	3-4	0.162	0.202	0.232	0.245	0.255	0.265	0.272
	5-9	0.288	0.337	0.374	0.389	0.405	0.414	0.422
	10-14	0.490	0.544	0.575	0.592	0.602	0.614	0.624
	15-19	0.595	0.654	0.687	0.700	0.710	0.717	0.723
	20-29	0.594	0.642	0.675	0.688	0.696	0.707	0.714
	30-39	0.677	0.725	0.747	0.759	0.766	0.774	0.781
	40-49	0.698	0.755	0.787	0.803	0.809	0.815	0.820
	50-99	0.815	0.854	0.877	0.889	0.894	0.897	0.902
	all	0.406	0.454	0.485	0.500	0.510	0.519	0.527
cd	3-4	0.197	0.257	0.301	0.318	0.332	0.344	0.354
	5-9	0.311	0.374	0.421	0.439	0.453	0.466	0.476
	10-14	0.486	0.538	0.574	0.588	0.599	0.608	0.616
	15-19	0.596	0.645	0.677	0.690	0.700	0.708	0.715
	20-29	0.609	0.660	0.695	0.707	0.717	0.726	0.733
	30-39	0.681	0.728	0.759	0.770	0.779	0.787	0.793
	40-49	0.707	0.758	0.789	0.800	0.808	0.815	0.821
	50-99	0.830	0.863	0.883	0.890	0.896	0.900	0.904
	all	0.423	0.479	0.518	0.533	0.545	0.555	0.563

Table continued...

	Number of quadrats species found in	Nest size (sq m)						
		4	9	16	20	24	28	32
acd	3-4	0.130	0.142	0.151	0.154	0.157	0.159	0.161
	5-9	0.257	0.281	0.296	0.302	0.307	0.311	0.314
	10-14	0.444	0.470	0.487	0.493	0.498	0.502	0.505
	15-19	0.543	0.565	0.579	0.584	0.588	0.591	0.594
	20-29	0.553	0.574	0.587	0.592	0.595	0.598	0.600
	30-39	0.626	0.644	0.655	0.659	0.662	0.664	0.666
	40-49	0.641	0.663	0.675	0.679	0.682	0.684	0.686
	50-99	0.772	0.787	0.795	0.797	0.799	0.801	0.802
	all	0.367	0.386	0.399	0.404	0.407	0.410	0.413
Mixed	3-4	0.172	0.214	0.243	0.254	0.263	0.270	0.277
	5-9	0.289	0.339	0.376	0.390	0.401	0.411	0.419
	10-14	0.469	0.512	0.541	0.552	0.561	0.568	0.575
	15-19	0.562	0.593	0.613	0.621	0.627	0.632	0.637
	20-29	0.572	0.606	0.629	0.638	0.644	0.650	0.655
	30-39	0.644	0.673	0.692	0.699	0.704	0.709	0.712
	40-49	0.658	0.690	0.709	0.716	0.721	0.725	0.728
	50-99	0.775	0.790	0.799	0.802	0.804	0.806	0.807
	all	0.395	0.435	0.463	0.473	0.481	0.488	0.494

Table 40 Ratio of predicted frequency to frequency of data by nest size and number of times species observed: Predictions from curves fitted to data from nests 1 m^2 and below

	Number of quadrats species found in	Nest size (sq m)						
		4	9	16	20	24	28	32
cd	3-4	1.21	1.27	1.30	1.30	1.30	1.30	1.30
	5-9	1.08	1.11	1.13	1.13	1.12	1.12	1.13
	10-14	0.99	0.99	1.00	0.99	0.99	0.99	0.99
	15-19	1.00	0.98	0.99	0.98	0.98	0.99	0.99
	20-29	1.03	1.03	1.03	1.03	1.03	1.03	1.03
	30-39	1.00	1.00	1.02	1.01	1.02	1.02	1.02
	40-49	1.01	1.00	1.00	1.00	1.00	1.00	1.00
	50-99	1.02	1.01	1.01	1.00	1.00	1.00	1.00
	all	1.04	1.06	1.07	1.07	1.07	1.07	1.07
acd	3-4	0.80	0.70	0.65	0.63	0.61	0.60	0.59
	5-9	0.89	0.83	0.79	0.78	0.76	0.75	0.74
	10-14	0.91	0.86	0.85	0.83	0.83	0.82	0.81
	15-19	0.91	0.86	0.84	0.83	0.83	0.82	0.82
	20-29	0.93	0.89	0.87	0.86	0.86	0.85	0.84
	30-39	0.92	0.89	0.88	0.87	0.86	0.86	0.85
	40-49	0.92	0.88	0.86	0.85	0.84	0.84	0.84
	50-99	0.95	0.92	0.91	0.90	0.89	0.89	0.89
	all	0.90	0.85	0.82	0.81	0.80	0.79	0.78
Mixed	3-4	1.06	1.06	1.05	1.04	1.03	1.02	1.02
	5-9	1.01	1.01	1.01	1.00	0.99	0.99	0.99
	10-14	0.96	0.94	0.94	0.93	0.93	0.93	0.92
	15-19	0.94	0.91	0.89	0.89	0.88	0.88	0.88
	20-29	0.96	0.94	0.93	0.93	0.93	0.92	0.92
	30-39	0.95	0.93	0.93	0.92	0.92	0.92	0.91
	40-49	0.94	0.91	0.90	0.89	0.89	0.89	0.89
	50-99	0.95	0.93	0.91	0.90	0.90	0.90	0.89
	all	0.97	0.96	0.95	0.95	0.94	0.94	0.94

Overview

7.33 The two methods of standardisation examined in this section are very different in nature. The first makes a simple adjustment to the frequencies of species in one dataset in order to make the average frequency the same as that of a second dataset. Though relatively crude the method can, in some situations, largely eliminate differences due to varying quadrat size. It has the advantage of being easy to apply, applicable to all datasets and not requiring nested data. However, it can, as shown for the CS dataset, eliminate real differences between datasets. A modification of the method equalises the average frequency of each site rather than of each dataset. The resulting analysis compares sites purely on the basis of species composition and not on the basis of composition and average frequency.

7.34 The second method of standardisation uses frequency-area curves fitted to individual species to predict the frequency of each species at the required quadrat size. Such a procedure is potentially superior to the previous crude frequency adjustment method since it should be capable of correctly estimating frequencies rather than just eliminating frequency differences. Achieving accurate prediction depends, however, on obtaining frequency-area curves that give accurate extrapolations outside the range of the observed data for all species concerned. In practise this is not achievable. In all three datasets there is insufficient data to fit a frequency-area curve to a substantial proportion of species and overall the average frequency of the predictions can differ substantially from the correct value. Furthermore the method is limited to datasets with information on a range of quadrat sizes, such as nested data.

8 Discussion, conclusions and recommendations

The effects of quadrat size

8.1 The primary purpose of this study was to investigate the utility of frequency-area curves in correcting for differences in quadrat size between vegetation datasets. However, the need for such corrections will only arise if varying quadrat size has a marked effect on the conclusions drawn from dataset comparisons. In Section 6 it was shown that ordinations obtained from different nest sizes within the same dataset are very similar suggesting that the same information and conclusions will be obtained regardless of which size is used. This was especially the case for the larger nest sizes, for which differences in the ordination plots were minimal, while smaller nest sizes were slightly more disparate, most probably as a result of the increased variation in plot composition at very small quadrat sizes. The implication is that variation in the form of frequency-area curves across species is limited. If curves varied markedly in their properties, crossing over each other, then different species would appear to be dominant at different scales, leading to differences in the corresponding ordinations. In Figure 85, for example, species A and B dominate at small quadrat sizes but species A and C at larger sizes. Such differences will arise as a result of the different spatial patterning of species, reflecting differences in dispersal methods, nutrient availability and other factors. The very similar results obtained for all nest sizes therefore suggests that such differences, at least as reflected in presence/absence data, are limited in their effect on community level data. This is an important finding since it confirms that choice of quadrat size is not crucial to individual studies.

8.2 Datasets compiled from mixed quadrat sizes, however, can give very different results from datasets in which all quadrats were the same size. In the mixed situation quadrat size tends to override the main axes of variation in the data and particular features of the data can be suppressed or exaggerated. Thus the study confirms that it is necessary to allow or correct for any quadrat size differences when making comparisons across datasets.

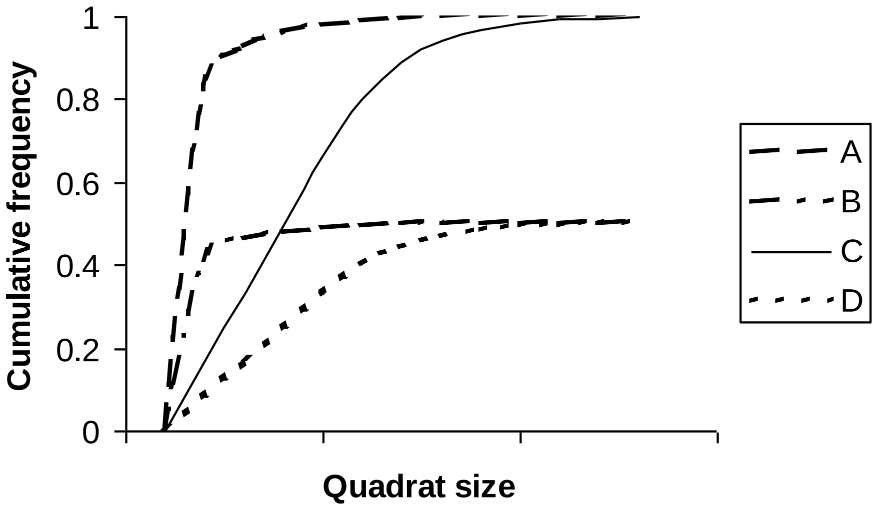


Figure 85 Illustration of curve differences

Choice and fitting of frequency-area curves

8.3 The range of functions considered here, based on the logistic and complementary log-log transformations, encompasses many of the functions commonly used in biological and environmental applications. It was clear from the analyses that the range was sufficiently wide to provide an adequate representation of the majority of the species studied for all of the datasets examined. The large number of analyses undertaken prevented detailed examination of the goodness of fit in each, however, so it is possible that in a small subset of cases none of the models considered adequately represented the data. Both types of model used in the study, the logistic and complementary log-log, are derived from linear regression applied after appropriate scale transformations (see Appendix 1). It would therefore be possible to extend the range of models by including quadratic, cubic or higher polynomial terms in the regression. This extended set of models would provide the basis for a more detailed examination of individual frequency-area curves in future work. The scope for extending the range of models is, however, limited by the number of nests since this puts an upper bound on the number of parameters that can be included in any model. The CS dataset with only six nest sizes is already close to being inadequate for fitting the most complex model used here which has four parameters. In addition the poor extrapolation properties of polynomial regression are well documented so the usefulness of such extended models for standardisation of datasets is questionable.

8.4 Given the differences between the datasets, and the variety and large number of species involved, it would not have been unexpected if each of the functions considered was found to be best for some subset of species. However there is clear evidence that just two of the parameterisations studied are capable of adequately representing the frequency-area relationships of the majority of species, one of which, the logistic *cd* parameterisation is a sub-model of the other, the logistic *acd*. Even more surprising, the same models were appropriate whether the data represented homogeneous vegetation, such as the individual NE sites, or very heterogeneous vegetation such as that from the complete CS dataset. Similar results were obtained from the logistic and complementary log-log models, suggesting that there is little difference in the ability of these two types of curve to represent observed frequency-area curves, though the logistic curves performed slightly better overall.

8.5 Fitting of the models is relatively straightforward though requiring specialist software and a moderate programming ability, partly because of the non-linear nature of the curves but essentially because of the nested data involved. The *cd* parameterisations, for example, could be fitted by most standard statistical packages if applied to data containing non-nested plots of a number of different sizes but require specialist software for fitting to nested data. It should be noted that treating nested data as if it were not nested will not give the optimum fitted frequency-area curve, though the extent of the inadequacy has not been investigated.

8.6 Convergence to optimum fitted values was rarely a problem. Where difficulties were found it was invariably due to inadequate data. To obtain accurate identification and representation of a frequency-area curve requires a reasonably large number of quadrats spread over a number of nest sizes. Ideally species to be modelled should be observed in at least twenty plots and at least six different nest sizes, preferably ten. The range of nest sizes is also important. Many species appear to exhibit a steep increase in frequency with area before reaching a relatively constant value. If possible the range of nest sizes should cover both parts of the curve.

8.7 Examination of the variation in fitted parameter values from the fitted logistic *acd* model showed that the fitted a parameters are essentially polarised into those having the value one and those having values close to the maximum observed frequency. Whilst this dichotomy could reflect a division of observed species into two types, those that occur over the whole of a studied site and those that are restricted to just part of it, it seems more likely to result from the use of nested data. Because nested data are constrained to increase in frequency with increasing nest size, most information on the a parameter is contained in the final or final few nest sizes, a property reflected in the constraining of the a parameter to be equal to or greater than the maximum observed frequency. Thus the dichotomy of the fitted a parameters throws doubt on the ability of data collected from nested quadrats to accurately estimate the value of this parameter. It also calls into question the utility of the *acd* parameterisation as a means of extrapolation and may explain the poorer predictive performance of the *acd* model described in Sections 7.27 – 7.32.

Ecological interpretation of frequency-area curve parameters

8.8 Understanding of the ecological/environmental meaning of the parameters of the fitted frequency-area curves is important if such curves are to be used as a research tool in their own right, though less important if their use is restricted to the standardisation of datasets. The most straightforward parameters to interpret are *a* and *d*. Figure 85 shows four curves that vary in terms of these parameters. Parameter a relates to the proportion of all plots occupied. A value of a less than one conveys partial occupancy of a site by a species and is equivalent to the occupancy measurements used in studies of local abundance versus range size (Gaston et al 1995). B and D are examples of such curves. Alternatively when a is one (curves A and C) then the curve approaches one with increasing nest size and the species is found in all plots of sufficiently large size.

8.9 The parameter *d* is inversely related to the average density of a species. Where the a parameter for a species is less than one d represents density in the occupied portion of the site. Curves A and B illustrate species with a small value of *d* resulting in a rapid increase in frequency with increasing plot size, modified in the case of curve B by the partial site occupancy. Curves C and D have a larger value of *d* resulting in more gentle increases in frequency as plot size increases.

8.10 Parameter *b* is a more difficult parameter to interpret ecologically though its mathematical interpretation is straightforward; it represents a lower limit to observed frequencies. Whereas, however, it is possible to interpret the upper limit a as a measure of site occupancy, a lower limit makes little ecological sense. The data would seem to agree with this statement since non-zero values of b only occur to any extent when the range of nest sizes analysed starts from a large quadrat size, that is, the analyses of the CS dataset and the ESA larger nest sizes. In these cases the non-zero values of *b* result from the fitting procedure taking advantage of the lack of constraining information at lower nest sizes to improve the fit to the larger nests, but at the cost of poorly representing behaviour at small, unobserved plot sizes.

8.11 The most difficult parameter to interpret is the *c* parameter. This parameter, mathematically a power transformation of the quadrat size scale, represents a measure of spatial patterning. When *c* takes the value one, the complementary log-log curve can be interpreted as representing complete spatial randomness in the distribution of a species and the logistic curve approximately the same. Values of *c* less than one represent a degree of spatial clustering of the species while values greater than one represent departure from randomness in the sense of a more regular or evenly spaced species distribution. It is interesting to see that for all three of the datasets used in this project median fitted values of *c* were less than one suggesting a moderate degree of spatial clustering for the majority of species.

Standardisation

8.12 Two methods of standardisation have been considered in this project. Unfortunately both are subject to problems. Standardisation using frequency adjustment enables comparison of the species composition of datasets to be made but will eliminate any differences in average frequency between datasets. Unfortunately there is no way to judge the extent of this loss of discrimination. If, however, differences in vegetation composition are thought to be of primary importance then this may not matter and the method has the advantage of being simple and straightforward to apply to any dataset.

8.13 The main problem with standardisation using frequency-area curves is the number of species that cannot be modelled accurately due to insufficient data. In Section 7 it was shown that curves fitted to species occurring in less than ten quadrats performed particularly poorly when used for prediction. Applying this criteria to the datasets used in this project 38%, 70% and 69% of the species records in the site specific NE, ADAS, and CS datasets respectively would not be accurately modelled. Rare species are always going to be a problem since increasing the number of quadrats will tend to find additional species that of necessity will be uncommon. However the proportion of species not meeting the criteria is surprisingly large. Furthermore it is not just rare species that may cause problems with accurate fitting. Very common species may occur in just a few nests making prediction at small nest sizes very inaccurate.

8.14 Thus frequency-area curves can only be used as a means of standardisation for part of a dataset. Some other method must be used for the rest of the data. An ad hoc procedure was used in the analyses reported here but this is unlikely to be satisfactory in many cases. A better alternative is to use a hybrid of the two standardisation methods, frequency-area curves for species with sufficient data and the frequency adjustment method for other species.

Conclusions and recommendations

8.15 In light of the results discussed above the main conclusion of this report must therefore be that in practise frequency-area curves have only a limited utility in standardising vegetation datasets. The reasons for this are:

1) Frequency-are curves can only be fitted to data from a variety of quadrat sizes, for example nested data, and can only therefore be used to compare two datasets if one of these is nested. They therefore do not provide the means to compare historic datasets with each other if these are collected using different non-nested quadrat sizes.
2) Accurate estimation of frequency-area curves can only be obtained for species with a substantial amount of data spread across a wide range of nest sizes. Many species can not therefore be modelled.
3) If data are available for a sufficient range of nest/quadrat sizes to enable accurate frequency-area curves to be estimated then it is likely that this range of nest sizes will cover or be sufficiently close to the nest size of the dataset with which comparisons need to be made (the target dataset), making standardisation redundant.
4) It was clear from the analyses performed that a proportion of species in each dataset were not accurately represented by the *cd* parameterisation, but also that nested data did not produce an accurate estimate, for extrapolation purposes, of the additional a parameter required.
5) The similarity of ordination diagrams obtained from different nest sizes suggests that differences between nest sizes are largely confined to differences in average frequency. Thus the need for adjustment on the individual species level is reduced.
6) Extrapolation using fitted frequency-area curves was shown to be poor with regard to predicted average frequency, especially for predictions at nest sizes less than 1 m^2.

8.16 If fitted frequency-area curves are to be used for standardisation the following procedure is recommended:

1) Restrict the species modelled to those occurring in more than 10 quadrats and 6 or more nest sizes.
2) Use the logistic *cd* curve for preference and the logistic *acd* curve only where the fit of the *cd* curve is clearly poor.
3) For species with insufficient data, scale the frequency they have in the nest closest in size to the target quadrat size by the ratio of the frequencies of the modelled species as predicted at the target nest size and in the nest closest to the target, restricting scaled frequencies to have values of no more than one.
4) If possible avoid standardisation to nest sizes less than 1 m^2 using data from nest sizes above 1 m^2.
5) If the average frequency in the standardised dataset is substantially different from the dataset to which it is to be compared then perform two separate analyses. The first with the datasets as they stand and the second after adjusting the frequencies in the dataset with the highest average value to have the same average frequency as the other dataset. If the results of the

two analyses differ then care should be taken in interpreting results. Average frequency in each case should be calculated over the combined list of species from both datasets, inserting zero frequency values where necessary.

6) If two nested datasets with no overlap in nest sizes are to be compared (for example the NE and CS datasets used in this study), make predictions from both datasets at a quadrat size intermediate between the two sets of nest sizes. This will minimise extrapolation errors.

8.17 Though of limited use for standardisation frequency-area curves may be of interest in their own right as a novel means of analysing nested quadrat data. The fitted parameters can themselves be analysed and the results of the analysis, when interpreted in terms of the ecological meaning of the parameters, may provide insights into differences between sites, vegetation types, species etc. that are not otherwise available. Such analyses are complemented by the use of ordination diagrams for all nest sizes, which indicate where different nest sizes give different results.

9 References

ADAS (1995). *ADAS National Strategy for ESA Monitoring*. (Unpublished report in 4 volumes).

AKAIKE, H. (1977). On entropy maximization principle. In Applications of Statistics, ed. P.R. Krishnaiah, pp. 27-41. North-Holland: Amsterdam.

AUSTIN, M.P. (1981). Permanent quadrats: an interface for theory and practice. *Vegetatio* 46: 1-10.

BARR, C.J., BUNCE, R.G.H., CLARKE, R.T., FULLER, R.M., FURSE, M.T., GILLESPIE, M.K., GROOM, G.B., HALLAM, CJ, HORNUNG, M., HOWARD, D.C. & NESS, M.J. 1993. *Countryside Survey 1990: Main Report. (Countryside Survey 1990 Vol. 2)*. Department of the Environment: London.

BUNCE, R.G.H. 1984. The use of simple data in the production of strategic sampling systems. In: BRANDT, J. and AGGER, P. (eds). *Methodology in landscape ecological research and planning. Vol. 4: Methodology of evaluation/synthesis of data in landscape ecology*. (1st Int. Seminar of the International Association of Landscape Ecology). Roskilde Univercity Centre: Roskilde.

BUNCE, R.G.H., BARR, C.J., GILLESPIE, M.K., HOWARD, D.C., SCOTT, W.A., SMART, S.M., VAN DE POLL, H.M. and WATKINS, J.W. (1999). *Vegetation of the British countryside – the Countryside Vegetation System*. ECOFACT Volume 1. Department of the Environment, Transport and the Regions: London.

COLLETT, D. (1991). *Modelling binary data*. Chapman and Hall: London.

CRITCHLEY, C. N. R. & POULTON, S. M. C. (1994). Biological monitoring of grasslands in Environmentally Sensitive Areas in England and Wales. In: *Grassland Management and Nature Conservation*, edited by HAGGAR, R. J. AND PEEL, S., 254–255. Occasional Symposium 28, British Grassland Society.

CRITCHLEY, C. N. R. & POULTON, S. M. C. (1998). A method to optimise precision and scale in vegetation monitoring and survey. *J. Veg. Sci*. 9(6): 837-846.

CRITCHLEY, C. N. R., POULTON, S. M. C., SMART, S. M. & MYERS, G. M. (1996a). Analysis procedures for ESA monitoring data (grasslands and allied habitats). (Unpublished ADAS report).

CRITCHLEY, C. N. R., SMART, S., POULTON, S. M. C. & MYERS, G. M. (1996). Monitoring the consequences of vegetation management in Environmentally Sensitive areas. *Aspects of Applied Biology* 44: 193–201.

CRITCHLEY, C. N. R. (1997). *Monitoring methods*. In: *Grassland Management in Environmentally Sensitive Areas*. BGS Occasional Symposium No. 32, [ed.] R. D. Sheldrick, pp. 44-54.

HAINES-YOUNG, R.H., BARR, C.J., BLACK, H.I.J., BRIGGS, D.J., BUNCE, R.G.H., CLARKE, R.T., COOPER, A., DAWSON, F.H., FIRBANK, L.G., FULLER, R.M., FURSE, M.T., GILLESPIE, M.K., HILL, R., HORNUNG, M., HOWARD, D.C., MCCANN, T., MORECROFT, M.D., PETIT, S., SIER, A.R.J., SMART, S.M., SMITH, G.M., STOTT, A.P., STUART, R.C., and WATKINS, J.W. (2000). Accounting for nature: assessing habitats in the UK countryside. Department of the Environment, Transport and the Regions: London.

HODGSON, J.G., COLASANTI, R. & SUTTON, F. (1995). *Monitoring grasslands: volume 1. English Nature Research Report 156*. Natural England: Peterborough.

Joint Nature Conservation Committee (1998). *A statement on common standards monitorin*g. JNCC: Peterborough.

MathSoft (1999). *S-Plus 2000 User's Guide*. Data Analysis Products Division, MathSoft: Seattle.

ROBERTSON, H. J., BINGHAM, J. & SLATER, I. (2000). Monitoring the condition of lowland grassland SSSIs. Part II: A test of the rapid assessment approach. English Nature Research Report 315. Natural England: Peterborough.

ROBERTSON, H. J. & JEFFERSON, R.G. (2000). *Monitoring the condition of lowland grassland SSSIs. Part I: Natural England's rapid assessment method. English Nature Research Report 315*. Natural England: Peterborough.

SAS Institute (1999). SAS/STAT *User's Guide, Version 8*. SAS Institute: Cary, NC.

Appendix 1 Curve types

Logistic

Derivation

This project is concerned with prediction of the frequency of occurrence of species within quadrats of different sizes. Throughout this report frequencies of occurrence are represented as proportions (that is, the proportion of quadrats in which particular species occur). Such data can only take values between 0 and 1. Statistically the most common way of dealing with data in the form of proportions is to transform them using the logistic transformation *z*=log(*y*/(*1*-*y*)). The purpose of this transformation is to convert the range of the data to a mathematically more tractable form since the transformed value, *z*, can take any real value whereas *y*, the frequency, is confined to the range 0 to 1.

Similarly quadrat sizes can only take positive values, since areas cannot be negative. Statistically the usual way of dealing with such data is to transform them using the log transformation *w*=log(*x*), where *x* is the quadrat area. This transformation has the same effect on positive values as the logistic transformation has on proportions; the range of the transformed value is no longer limited.

Again the standard statistical technique for studying the relationship between two variables is regression. If it is assumed that the relationship between the transformed frequencies, *z*, and the transformed quadrat areas, *w*, can be represented by the simple linear regression;

z=cw+D

then replacing the transformed variables by their representations in terms of the original variables gives;

log(y/(1-y))=c*log(x)+D

or;

y=xc/(xc+d)

where d=exp(-D).

Generalised Logistic

Derivation

Suppose now that the frequencies, y, do not vary from 0 to 1 but are instead limited to the range *b* to *a*. Application of the logistic transform will no longer produce a variable of unlimited range. However the preliminary transformation;

y=(y-b)/(a-b)*

produces a variable, *y**, which ranges from 0 to 1. Applying the logistic transform to this rescaled variable and regressing on log(*x*) gives, after some reorganisation, the generalised logistic equation;

$$y = a\left(\frac{x^c + bd}{x^c + d}\right)$$

This four-parameter curve provides a flexible means of representing frequency-area relationships with the added benefit that all of the parameters can be interpreted in ecologically meaningful terms. The parameter a represents an upper limit to frequencies, *b* represents a lower limit, *c* provides a transformation of the area scale, and *d* determines the rate at which frequency of occurrence increases with area.

Complementary log-log

Derivation

An alternative transformation to the logistic for data in the form of proportions is the complementary log-log transformation *z*=log(-log(*1-y*)). As with the logistic, the purpose of this transformation is to convert the range of the data so that the transformed value, *z*, can take any real value.

Regressing the transformed variable on log (quadrat size) as above gives, in terms of the original variables;

log(-log(1-y))=c*log(x)+D

or;

y=1-exp(-dxc)

where d=exp(-D).

As with the logistic this equation can be generalised to incorporate limits on the range of *y* other than 0 to 1. Allowing *y* to vary from *b* to *a* gives the equation;

$$y = a - (a - b)\exp(-dx^c)$$

an alternative four parameter curve. The parameters have the same interpretation as for the logistic.

Relationship between curves

Fixing the parameters of either the generalised logistic curve or the generalised complementary log-log curve at suitable values generates a family of curves that can be used to model frequency-area relationships. The obvious fixed values to use are:

- a=1, representing no upper limit on frequency values
- b=0, representing no lower limit on frequency values
- c=1, representing no transformation of quadrat area
- d=1, representing zero intercept in the regression on log (quadrat size).

Whilst fixing any of the first three parameters in this way can be interpreted in an ecologically meaningful way, there appears to be little obvious justification for fixing the fourth parameter. Consequently this option will not be considered further.

The inter-relationships of the curves generated by fixing combinations of the first three parameters can be shown diagrammatically. In Figure A each box represents a particular curve with its parameters written in the box. Parameters not included in a box are fixed at the values listed above. The curves are arranged hierarchically according to the number of parameters with the highest number of parameters at the top. Lines are drawn between curves that can be compared using formal statistical tests (see Appendix 2) and such tests correspond to testing whether parameters present in one curve but not another are needed to describe the data.

In total therefore fourteen different curves are examined in this project, the seven curves represented in Figure A for each of the logistic and complementary log-log transformations. Some of these curves have particular interpretations. For example the complementary log-log curve containing just parameter *d* represents a model of spatial randomness. Species following this curve would be distributed across a site completely at random. The logistic curve with parameters *a* and *d* is the curve fitted by Hodgson et al. (1995) to nested quadrat data and called by them a two parameter hyperbola.

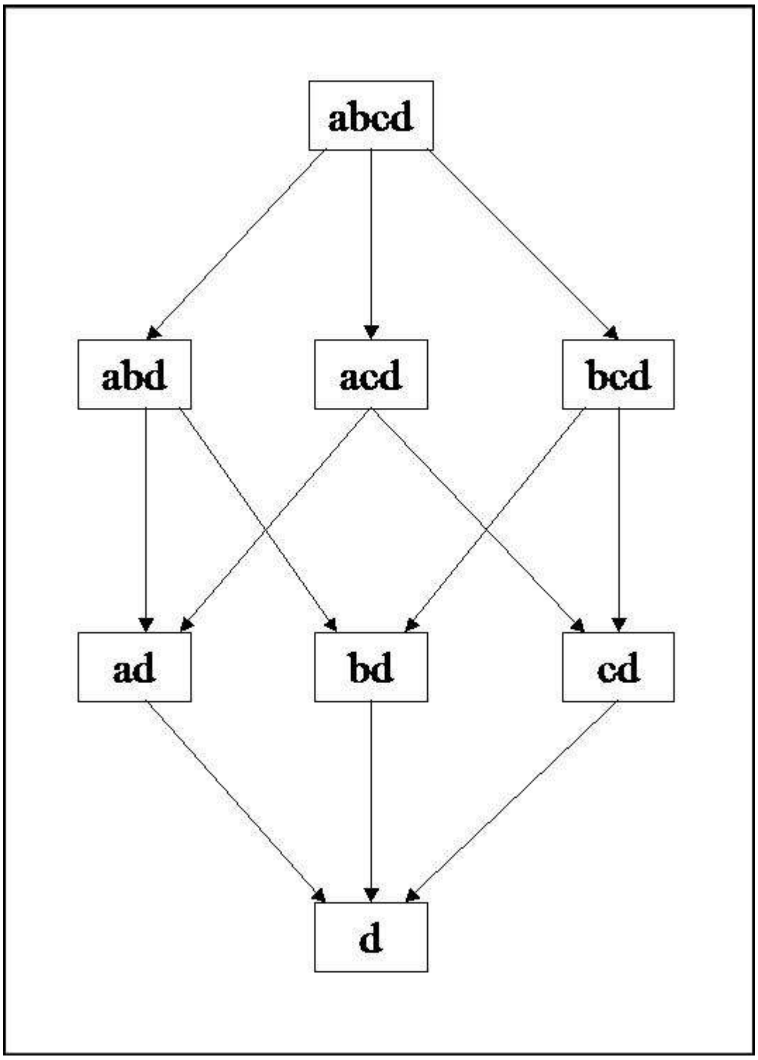


Figure A Representation of relationships between curves

Appendix 2 Curve fitting and model testing

Nested quadrats

Previous studies involving data from nested quadrats have usually fitted curves by regression. Such an approach is incorrect since it treats the data from each nest size as if it were independent. A more statistically valid approach is to use maximum likelihood methods. Let the frequency-area curve for a particular species of interest be represented by the equation;

$$y = p(x \mid \theta)$$

where *y* is frequency, *x* is quadrat size, *θ* represents the set of curve parameters and *p* is one of the expressions given in Table 4 for converting quadrat size to frequency. Suppose that information is available for a set of *N* nested quadrats with *K* levels of nesting. Let *ki* represent the nest level at which the species was found in quadrat *i*, where *ki* can take values from *1* to *K+1*, a value of *K+1* indicating that the species was not found in quadrat *i*. Let *x(k)* represent the nest size at level *k* and *p(k)=p(x(k)|θ* be the frequency at nest size *x(k)* with *p(0)=0* and *p(K+1)=1*. Then the probability that the species is first found in quadrat *i* at nest level *ki* can be shown to be *p(ki)-p(ki-1)*. The likelihood over the complete set of quadrats is therefore;

$$L = \prod_{i=1}^{N} \{ p(k_i) - p(k_i - 1) \}$$

a function of the curve parameters, *θ.* Finding the best fitting curve for a specific species is then equivalent to finding the value of *θ* that maximises the value of *L*.

Non-nested quadrats

Though all data used in this study was nested, estimation methods for non-nested quadrats are included here for completeness. Suppose now that a dataset is available which contains species information from quadrats of different sizes. Using the same notation as above and letting *xi* represent the size of quadrat *i* the likelihood for a set of N quadrats is;

$$L = \prod_{i=1}^{N} p(x_i)$$

and this can be maximised as before to give the parameters of the best fitting curve. It is difficult to imagine, however, a situation in which sufficient different quadrat sizes would have been used to make this technique of practical value.

Model comparison

Maximising the likelihood as described above determines the best fitting model of a particular type, but it is also desirable to be able to assess whether one type of model is a better fit than another. Figure A shows the relationships between models and indicates which models can be formally compared. For any two such models the deviance, defined as twice the difference in the log of the likelihoods of the two models, can be shown to have a chi-square distribution with degrees of freedom equal to the difference in the number of parameters providing statistical tests and levels of significance. Curves not directly connected can be compared informally using the deviance of the fitted models or the Akaike information criteria, a modification of the deviance to take account of differences in the number of parameters (Akaike, 1977).

Natural England works for people, places and nature to conserve and enhance biodiversity, landscapes and wildlife in rural, urban, coastal and marine areas.

www.naturalengland.org.uk